4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), April 2003, Nara, Japan

Advances in Nonlinear Blind Source Separation

Christian Jutten
Lab. des Images et des Signaux, INPG
46, avenue Felix Viallet
8031 Grenoble Cedex, FRANCE
Email: Christian.Jutten@inpg.fr

Juha Karhunen
Neural Networks Research Centre
Helsinki University of Technology
P.O. Box 5400, 02015 HUT, FINLAND
Email: Juha.Karhunen@hut.fi

***Abstract*— In this paper, we briefly review recent advances in blind source separation (BSS) for nonlinear mixing models. After a general introduction to the nonlinear BSS and ICA (independent Component Analysis) problems, we discuss in more detail 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 extra constraints, which are often implemented by using a suitable regularization. Post-nonlinear mixtures are an important special case, where a nonlinearity is applied to linear mixtures. For such mixtures, the ambiguities are essentially the same as for the linear ICA or BSS problems. In the later part of this paper, various separation techniques proposed for post-nonlinear mixtures and general nonlinear mixtures are reviewed.**

I. The nonlinear ICA and BSS problems

Consider N samples of the observed data vector $\mathbf{x}$, modeled by

$$\mathbf{x} = \mathbf{A}\mathbf{s} = \sum_{j=1}^{n} s_j \mathbf{a}_j \tag{1}$$

where $\mathbf{A}$ is the unknown mixing matrix with column vectors $\mathbf{a}_j$, $j = 1, 2, \ldots, n$, and $\mathbf{s}$ 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. In general, the dimensionality m of the vectors $\mathbf{x}$ and $\mathbf{a}_j$ can be different from n. Usually it is assumed that there are at least as many mixtures as sources ($m \geq n$), the mixing matrix $\mathbf{A}$ has full rank, and that at most one of the sources s_j is Gaussian.

ICA is a method which consists in estimating a matrix $\mathbf{B}$ such that $\mathbf{y} = \mathbf{B}\mathbf{x}$ are statistically independent, only from the observed data $\mathbf{x}$.

Such a model can be used in different situations, for example:

- In multidimensional signal processing, where each sensor receives an unknown superimposition of unknown source signals at time instants $t = 1, \ldots, N$.
- In sparse coding, where one tries to code the N data vectors as a (sparse) linear combination of independent components.

In the first problem, 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. The second problem is related to a hypothetical and rough model which tries to approximate the data as well as possible using n suitably chosen independent components [1].

In both cases, since the only assumption is the independence of sources, ICA is used for solving the problems. It has been proved that ICA and BSS are equivalent (with well-known indeterminacies) with the above assumptions [2], and this basic (linear) case is now understood quite well [3], [1], [4]. Since 1985 (see [5] for a historical review and early references), several well-performing BSS and ICA algorithms [6], [7], [8], [9], [10], [11], [12] have been developed and applied to an increasing number of applications [13], [14], [15], [16], [17], [1], [18], [4]. Many more references on linear ICA and BSS can be found in the recent books [1], [4].

If the source signals are not plain random variables but have a temporal structure, linear blind source separation can be achieved by utilizing either temporal correlations [19], [20] or nonstationarity [21], [22] under somewhat different assumptions; see [1], Chapter 18. Moreover, the basic model (1) is often too simple for describing the observed data $\mathbf{x}$ adequately.

A natural extension of the linear model (1) is to consider nonlinear mixing models. For instantaneous mixtures, the nonlinear mixing model has the general form

$$\mathbf{x} = \mathcal{F}(\mathbf{s}) \tag{2}$$

where $\mathbf{x}$ and $\mathbf{s}$ denote the data and source vectors as before, and $\mathcal{F}$ is an unknown real-valued m-component mixing function.

Assume now for simplicity that the number of independent components n equals the number of mixtures m. The general nonlinear ICA problem then consists of finding a mapping $\mathcal{G} : \mathcal{R}^n \to \mathcal{R}^n$ that yields components

$$\mathbf{y} = \mathcal{G}(\mathbf{x}) \tag{3}$$

which are statistically independent. A fundamental characteristic of the nonlinear ICA problem is that in the general case, solutions always exist, and they are highly nonunique. One reason for this is that if x and y are two independent random variables, any of their functions $f(x)$ and $g(y)$ are also independent. An even more serious problem is that in the nonlinear case, x and y can be mixed and still be statistically independent (see Section II).

Contrary to the linear case, the BSS problem for general nonlinear mixtures differs greatly from the nonlinear ICA problem defined above. In the respective nonlinear BSS problem, one should find the original source signals $\mathbf{s}$ that have generated the observed data $\mathbf{x}$. This is usually a clearly more meaningful and unique problem than the nonlinear ICA problem defined above, provided that suitable prior information is available on the sources and/or the mixing mapping. If some arbitrary independent components are found for the data generated by (2), they may be quite different from the true source signals. Generally, solving the nonlinear BSS problem is not easy, and requires additional prior information or suitable regularizing constraints.

II. EXISTENCE AND UNIQUENESS OF NONLINEAR ICA AND BSS

Several authors [23], [24], [5], [25], [26] 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 [27], are reported in this section.

A. Indeterminacies

Recall first the definition of independent random vector.

Definition 1.1: A random vector $\mathbf{x}$ is statistically independent 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.

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

Definition 1.2: A one-to-one mapping $\mathcal{H}$ is called *trivial*, if it transforms *any* random vector $\mathbf{s}$ with independent components into a random vector with independent components.

The set of trivial transformations will be denoted by $\mathfrak{T}$. Trivial mappings preserve the independence property of *any* random vector. One can easily show that a one-to-one mapping $\mathcal{H}$ is trivial if and only if it satisfies

$$\mathcal{H}_i(u_1, u_2, \ldots, u_n) = h_i(u_{\sigma(i)}),\ i = 1, 2, \ldots, n \qquad (4)$$

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

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 transformation $\mathcal{H} = \mathcal{G} \circ \mathcal{F}$ trivial using the independence assumption.

However, from (4) it is clear that sources can only be separated up to a permutation and a nonlinear function. For any invertible mapping $\mathcal{F}(\mathbf{x}) = [f_1(\mathbf{x}), \ldots, f_n(\mathbf{x})]^T$ whose each component 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 $p_{\mathcal{F}(\mathbf{x})}(\mathbf{v}) = \prod_i p_{f_i(x_i)}(v_i)$. Moreover, this is not possible without imposing additional constraints on $\mathcal{H}$, as we shall see in the next subsection.

B. Results from factor analysis

In the general case when the mapping $\mathcal{H}$ has no particular form, a well-known statistical result shows that preserving independence is not a strong enough constraint for ensuring tseparability in the sense of equation (4). This result has been established already early in the 50's by Darmois [27]. He used a simple constructive method, similar to the well-known Gram-Schmidt orthogonalization procedure in linear algebra, for decomposing any random vector into a non-trivial mapping of independent variables. In [23], Hyvärinen and Pajunen have applied Darmois's idea to construction of parameterized families of nonlinear ICA solutions.

Darmois's result is negative in the sense that it shows that there exist non-trivial transformations $\mathcal{H}$ which *"mix"* 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.

In the conclusion of [27], Darmois clearly states: *"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.*"

C. A simple example

We give a simple example of mixing mappings preserving independence, derived from [24]. Suppose $s_1 \in \mathcal{R}^+$ 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, having a uniform pdf $s_2 \in [0, 2\pi)$. Consider the nonlinear mapping

$$\begin{aligned} [y_1, y_2] &= \mathcal{H}(s_1, s_2) \\ &= [s_1 \cos(s_2), s_1 \sin(s_2)] \end{aligned} \qquad (5)$$

which has a non-diagonal Jacobian matrix

$$\mathbf{J} = \begin{pmatrix} \cos(s_2) & -s_1 \sin(s_2) \\ \sin(s_2) & s_1 \cos(s_2) \end{pmatrix} \qquad (6)$$

The joint pdf of y_1 and y_2 is

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

This shows that the random variables y_1 and y_2 are independent, although they are still nonlinear mixtures of the sources.

Other examples can be found in the literature (see for example [28]), or can be easily constructed.

D. Specific model

The basic reason for Darmois's negative result is that no constraints were assumed on the transformation $\mathcal{H}$. Constraining the transformation $\mathcal{H}$ in a certain set of transformations $\mathfrak{Q}$ can reduce these great indeterminacies.

1) Smooth mappings: Recently, multi-layer perceptron (MLP) networks (see [29]) have been used in [30] for estimating the generic nonlinear mappings $\mathcal{H}$. It is conjectured that smooth mappings providing by MLP networks are sufficient for ensuring that nonlinear ICA leads to nonlinear BSS, too. However, the following example [31] shows that smoothness alone is not sufficient for separation.

Without a loss of generality, consider two independent random variables $\mathbf{x} = (x_1, x_2)^T$ which are both uniformly distributed in the interval $[-1, 1]$, and the nonlinear smooth mapping

$$\mathbf{R} = \begin{pmatrix} \cos(\theta(r)) & -\sin(\theta(r)) \\ \sin(\theta(r)) & \cos(\theta(r)) \end{pmatrix} \tag{7}$$

This is a rotation for which the rotation angle $\theta(r)$ depends on the radius $r = (x_1^2 + x_2^2)^{1/2}$ as follows:

$$\theta(r) = \begin{cases} \theta_0(1-r)^q, & 0 \leq r \leq 1 \\ 0, & r > 1 \end{cases} \tag{8}$$

where $q \geq 2$. This smooth mapping $\mathbf{R}$ preserves independence since the Jacobian of the transformation is equal to 1, but it is still mixing since the Jacobian matrix is not diagonal. This counterexample proves that restricting the mapping to be smooth is not sufficient.

Since smoothness is too vague, one has to explore further for defining sufficient conditions. Hyvärinen and Pajunen gave a partial answer to this question in [23], proving that a unique solution (up to a rotation) can be obtained in the two-dimensional special case if the mixing mapping $\mathcal{F}$ is constrained to be a conformal mapping.

2) 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

$$\forall E \in \mathfrak{M}_n$$
$$\int_E dF_{s_1} dF_{s_2} \cdots dF_{s_n} = \int_{\mathcal{H}(E)} dF_{y_1} dF_{y_2} \cdots dF_{y_n} \tag{9}$$

where $\mathfrak{M}_n$ is a σ-algebra on $\mathbb{R}^n$.

Let $\mathfrak{P}$ denote the set[1]

$$\mathfrak{P} = \{(F_{s_1}, F_{s_2}, \ldots, F_{s_n}), / \exists \mathcal{H} \in \mathfrak{Q} \setminus (\mathfrak{T} \cap \mathfrak{Q}) : \mathcal{H}(\boldsymbol{s}) \text{ has independent components}\} \tag{10}$$

of all source distributions $(F_{s_1}, F_{s_2}, \ldots, F_{s_n})$ for which there exists a non-trivial mapping $\mathcal{H}$ 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 $\mathfrak{T} \cap \mathfrak{Q}$.

[1] In equation (10), $\setminus$ denotes the difference between two sets

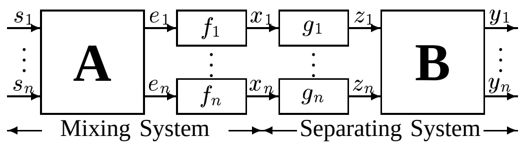

Fig. 1. The mixing-separating system for PNL mixtures.

E. Example: Linear models

In the case of regular linear models, the transformation $\mathcal{F}$ is linear and can be represented by (1), where $\mathbf{A}$ is a square invertible matrix. In this case it suffices to constrain the separating model $\mathcal{G}$ 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 $\mathcal{H}$ 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 matrix. From the Darmois-Skitovich theorem [27], it is clear that the set $\mathfrak{P}$ contains the distributions $(F_{s_1}, F_{s_2}, \ldots, F_{s_n})$ such that at least two sources, *e.g.* F_{s_i} and F_{s_j}, are Gaussian. Thus we end up with Comon's well-known theorem [2]: 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.

F. Separability of PNL mixtures

In the post-nonlinear (PNL) model, the nonlinear observations have the the following specific form (Figure 1):

$$x_i(t) = f_i(\sum_{j=1}^{n} a_{ij} s_j(t)), \; i = 1, \ldots, n \tag{11}$$

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[2] family suits perfectly for a lot of real-world applications. For instance, such models appear in sensors array processing [32], satellite and microwave communications [33], and in many biological systems [34].

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

1) It can invert the mixing system in the sense of Eq. (4).
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 $\mathcal{G}$ the mirror structure of the mixing system $\mathcal{F}$, see Figure 1. The global transform $\mathcal{H}$ is

[2] L stands for Linear and ZMNL stands for Zero-Memory NonLinearity.

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 [24], it has been shown that these mixtures are separable for distributions having at most two Gaussian sources (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 matrix) if $\mathbf{A}$ has at least 2 nonzero entries on each row and column.

Separability of PNL mixtures can be generalized to convolutive PNL mixtures, in which the instantaneous mixtures (matrix $\mathbf{A}$) is replaced by linear filters (matrix of filters $\mathbf{A}(z)$ [35]. Using a suitable parameterization, Wiener systems can be viewed as particular PNL mixtures. Consequently, separability of PNL mixtures ensures blind invertibility of Wiener systems [36]. Theis et al. have studied in [37] separability of a cascade of PNL stages, constituting a structure similar to multi-layer perceptron networks.

G. 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.

1) A simple example: As an example, consider multiplicative mixtures:

$$x_j(t) = \prod_{i=1}^{n} s_i^{\alpha_i}(t),\ j = 1, \ldots, n \tag{12}$$

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

$$\ln x_j(t) = \sum_{i=1}^{n} \alpha_i \ln s_i(t),\ j = 1, \ldots, n \tag{13}$$

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 [38], or gray-level images as a product of incident light and reflected light [26]. Considering in more detail the former example, the Hall voltage [39] is equal to

$$V_H = kBT^{\alpha} \tag{14}$$

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

$$\left\{ \begin{array}{lcl} V_{H_N}(t) & = & k_N B(t) T^{\alpha_N}(t) \\ V_{H_P}(t) & = & k_P B(t) T^{\alpha_P}(t) \end{array} \right. \tag{15}$$

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

$$\left\{ \begin{array}{lcl} \ln \mid V_{H_N} \mid & = & \ln k_N + \ln \mid B \mid + \alpha_N \ln T \\ \ln \mid V_{H_P} \mid & = & \ln k_P + \ln \mid B \mid + \alpha_P \ln T \end{array} \right. \tag{16}$$

These equations describe a linear mixture of the two sources $\ln \mid B \mid$ 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:

$$R = \frac{V_{H_N}}{V_{H_P}} = \frac{k_N}{k_P} T^{\alpha_N - \alpha_P} \tag{17}$$

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.

2) Generalization to a class of mappings: Extension of the Darmois-Skitovic theorem to nonlinear functions has been addressed by Kagan et al. in [40]. Their results have recently been revisited within the framework of BSS of nonlinear mixtures by Eriksson and Koivunen [26]. The main idea is to consider particular mappings $\mathcal{F}$ 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:

$$\left\{ \begin{array}{lcl} x_1 & = & (s_1 + s_2)(1 + s_1 s_2)^{-1} \\ x_2 & = & (s_1 - s_2)(1 - s_1 s_2)^{-1} \end{array} \right.$$

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

$$\left\{ \begin{array}{lcl} x_1 & = & \tan(u_1 + u_2) \\ x_2 & = & \tan(u_1 - u_2) \end{array} \right.$$

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

$$\left\{ \begin{array}{lcl} v_1 & = & \tan^{-1}(x_1) = u_1 + u_2 \\ v_2 & = & \tan^{-1}(x_2) = u_1 - u_2 \end{array} \right.$$

which is now a linear mixture of the two independent variables u_1 and u_2. This nice result is due to the 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 $\mathcal{F}$ and an invertible function f satisfying an addition theorem:

$$f(s_1 + s_2) = \mathcal{F}[f(s_1), f(s_2)] \tag{21}$$

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

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

In other words, denoting $u \circ v = \mathcal{F}(u, v)$, these conditions imply that the set $(\mathfrak{G}, \circ)$ is an Abelian group. Under this condition, Aczel [41] proved that there exists a monotonic and continuous function $f : \mathbb{R} \to [a, b]$ such that

$$f(x + y) = \mathcal{F}(f(x), f(y)) = f(x) \circ f(y) \qquad (22)$$

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

$$x + y = f^{-1}(\mathcal{F}(f(x), f(y))) = f^{-1}(f(x) \circ f(y)) \qquad (23)$$

Using the above property (22), one can define a product $\star$ with integer and extend it to real variables:

$$f(cx) = c \star f(x) \qquad (24)$$

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

$$cf^{-1}(u) = f^{-1}(c \star u) \qquad (25)$$

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

$$c_1 f^{-1}(u_1) + \ldots c_n f^{-1}(u_n) = f^{-1}(c_1 \star u_1 \circ \ldots \circ c_n \star u_n) \quad (26)$$

Finally, Kagan et al. [40] stated the following theorem:

Theorem 7.1: Let $u_1, \ldots, u_n$ be independent random variables such that

$$\begin{cases} x_1 &= a_1 \star u_1 \circ \ldots \circ a_n \star u_n \\ x_2 &= b_1 \star u_1 \circ \ldots \circ b_n \star u_n \end{cases}$$

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 $a_i b_i \neq 0$.

This theorem can be easily extended to source separation, and with such mixtures the separation algorithm consists of 3 practical steps [26]:

- Apply f^{-1} to the nonlinear observations for providing linear mixtures in $s_i = f^{-1}(u_i)$.
- Solve the linear mixtures in s_i by any BSS method.
- 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.* [24]) can be used for *blindly* separating these nonlinear mappings, avoiding the above step 1 used in [26]. Other examples of mappings satisfying the addition theorem are given in [40], [26]. However, realistic mixtures belonging to this class seem unusual, except for the PNL mixtures (11) and the multiplicative mixtures (12).

Taleb and Jutten have considered separability of nonlinear mixtures in [24], [5]. Their general conclusion is the same as earlier: Separation is impossible without additional prior knowledge on the model, since the independence assumption alone is not strong enough in the general nonlinear case.

H. Prior information on the sources

In this subsection we show that prior information on the sources can simplify or relax the indeterminacies in nonlinear mixtures.

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 random sample of a sine wave satisfy this condition. For simplicity, we discuss only PNL mixtures (Figure 1) of two sources, but the results can be easily extended to more sources. The joint distribution of the two sources $\mathbf{s}$ is then contained in a rectangle. After the linear mixing $\mathbf{As}$, the joint distribution of $\mathbf{e}$ 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.

Babaie-Zadeh, Jutten and Nayebi [42] proved that a distribution contained in a parallelogram (DCP) can be transformed by componentwise invertible mappings into another DCP only if the mappings are linear. The nonlinearities g_i compensating f_i can then be estimated so that the borders of the distorted parallelogram associated with the joint distribution of the PNL mixtures become straight lines. Details of the algorithm and experimental results are given in [42]. This method proves that 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.

2) Time correlated sources in nonlinear mixtures: Consider two independent and identically distributed random signals, $s_1(t)$ and $s_2(t)$. Using the Darmois decomposition procedure [27], [23], one can construct new signals $y_1(t)$ and $y_2(t)$ which are statistically independent although the underlying mapping is still a mixing mapping:

$$\begin{cases} y_1(t) &= F_{X_1}(s_1(t)) \\ y_2(t) &= F_{X_1|X_2}(s_1(t), s_2(t)) \end{cases}$$

Here F_X denotes the cumulative probability function of the random variable X. If the sources are temporally correlated, Hosseini and Jutten proved [43] that the above mapping does no longer preserve independence. Of course, this partial 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 information on the sources can reduce the typical indeterminacies of ICA encountered in nonlinear mixtures.

III. Separation methods for post-nonlinear mixtures

A. Minimization of mutual information

Consider now BSS methods proposed for the simpler case of post-nonlinear mixtures (11). Taleb and Jutten have studied this case in several papers [44], [24], [45], and we start with a brief discussion of their results. A short overview of their studies can be found in [5], and the main results have been represented in [24].

The separation algorithm for the post-nonlinear mixtures (11) generally consists of two subsequent parts or stages:

1) A *nonlinear stage*, which should cancel the nonlinear distortions f_i, $i = 1, \ldots, n$. This part consists of nonlinear functions $g_i(\boldsymbol{\theta}_i, u)$.
2) A *linear stage* that separates the approximately linear mixtures $\mathbf{z}$ obtained after the nonlinear stage. This is done as usual by learning an $n \times n$ separating matrix $\mathbf{B}$ for which the components of the output vector $\mathbf{y} = \mathbf{Bz}$ of the separating system are statistically independent (or as independent as possible).

Taleb and Jutten [24] use the mutual information $I(\mathbf{y})$ between the components $y_1, \ldots, y_n$ of the output vector as the cost function and independence criterion in both stages. For the linear part, minimization of the mutual information leads to the same estimation equations as for linear mixtures [1], [4]

$$\frac{\partial I(\mathbf{y})}{\partial \mathbf{B}} = -\mathrm{E}\{\boldsymbol{\psi}\mathbf{x}^T\} - (\mathbf{B}^T)^{-1} \tag{29}$$

where components ψ_i of the vector $\boldsymbol{\psi}$ are score functions of the components y_i of the output vector $\mathbf{y}$:

$$\psi_i(u) = \frac{d}{du}\log p_i(u) = \frac{p_i'(u)}{p_i(u)} \tag{30}$$

Here $p_i(u)$ is the pdf of y_i and $p_i'(u)$ its derivative. In practice, the natural gradient algorithm [8], [10], [11] is used for providing equivariant performance, which does not depend on the mixing matrix $\mathbf{A}$ provided that there is no noise present.

For the nonlinear stage, one can derive from the estimating equations the gradient learning rule [24]

$$\begin{aligned}\frac{\partial I(\mathbf{y})}{\partial \boldsymbol{\theta}_k} = &-\mathrm{E}\left\{\frac{\partial \log \mid g_k'(\boldsymbol{\theta}_k, x_k) \mid}{\partial \boldsymbol{\theta}_k}\right\} \\ &- \mathrm{E}\left\{\sum_{i=1}^{n} \psi_i(y_i) b_{ik} \frac{\partial g_k(\boldsymbol{\theta}_k, x_k)}{\partial \boldsymbol{\theta}_k}\right\}\end{aligned} \tag{31}$$

Here x_k is the kth component of the observation vector, b_{ik} is the element ik of the separating matrix $\mathbf{B}$, and g_k' is the derivative of the kth nonlinear function g_k. The exact computation algorithm depends naturally on the specific parametric form of the nonlinear mapping $g_k(\boldsymbol{\theta}_k, x_k)$. In [24], a multilayer perceptron network is used for modeling the functions $g_k(\boldsymbol{\theta}_k, x_k)$, $k = 1, \ldots, n$.

Contrary to BSS of linear mixtures, separation performance for nonlinear mixtures is strongly related to the estimation accuracy of the score functions (30) [24]. The score functions (30) must be estimated adaptively from the output vector $\mathbf{y}$. Several alternative ways to do this are considered in [24]. The first approach is to estimate the pdf, and then compute using differentiation the score function. Pdf estimation based on the Gram-Charlier expansion [2], [1] performs appropriately only for mild post-nonlinear distortions. For hard nonlinearities, a simple pdf estimation based on kernel methods is preferable. The second method estimates the score functions directly, and provides very good results for hard nonlinearities, too. A well performing batch type method for estimating the score functions has been introduced in a later paper [45].

B. Other methods for post-nonlinear mixtures

Several other authors have studied methods for blind separation of post-nonlinear mixtures starting from different viewpoints. An early method proposed by Lee, Koehler, and Orglmeister in [46] is based on an extension of the natural gradient algorithm, and uses either parametric sigmoidal nonlinearities or more flexible higher-order polynomials. Achard, Pham and Jutten [47] proposed another parameterization of the mutual information criterion involving the derivatives of the nonlinearities. As these derivatives are parameterized by piecewise constant functions, the algorithms are very simple.

Another approach by Ziehe et al. [48] uses first the alternating conditional expectation (ACE) method of nonparametric statistics for approximate inversion of the post-nonlinearities f_i in (11). After this a BSS method called TDSEP, based on temporal decorrelation and introduced earlier by the same authors (see [48]), is used for recovering the source signals. Independently, Solé et al. [49] and Ziehe et al. [50] improved the method by directly computing (instead of estimating) the inverse g_i (see Figure 1) of the nonlinear mapping f according to the formula

$$\hat{g}_i = \Phi^{-1} \circ F_{X_i} \tag{32}$$

Here F_{X_i} is the cumulative distribution function of the random variable X_i, and Φ is the cumulative Gaussian distribution.

Peng, Chi, and Siu [51] have introduced a semi-parametric hybrid neural network model based on the MLP network [29] for separating post-nonlinear mixtures. The main advantage of their method is that it is able to consider cross-channel post-nonlinearities, too, but experimental results have been presented on separation of two sources only. A similar generalized post-nonlinear mixture model has been addressed also in [52] by using adaptive spline neural networks.

In [53], Puntonet et al. have proposed combining a geometric approach with neural network learning for separating a special class of post-nonlinear mixtures, which are assumed to be powers of linear mixtures of the source signals. Another geometrical approach for separating bounded sources in PNL mixtures has been proposed by Babaie-Zadeh, Jutten and Nayebi in [42].

C. Extension of PNL mixtures

A Wiener system consists of the cascade of a linear filter $[H(z)]$ followed by a memoryless nonlinearity f, whose input is an independent and identically distributed signal $s(k)$. The output is then $x(k) = f([H(z)]s(k))$. Using a suitably chosen parameterization, Taleb, Solé and Jutten proved that Wiener systems can be expressed as PNL mixtures, and proposed non parametric [36] as well as parametric [54] algorithms based on minimization of mutual information rate [55]. A similar problem appearing in satellite communications has been solved using Monte Carlo Markov Chain (MCMC) simulation methods [56].

Convolutive post-nonlinear (CPNL) mixtures have been introduced by Babaie-Zadeh, Jutten and Nayebi for taking into account propagation which is commonplace in many realistic situations. The observation vector is then

$$x_i(k) = f_i([A(z)]s(k)),\ i = 1, \ldots, n \tag{33}$$

Some separation algorithms based on the generalization of mutual information minimization for random processes have been proposed in [35], [57].

IV. Separation methods for general nonlinear mixtures

A. Variational Bayesian methods

Advanced Bayesian inference methods are becoming increasingly popular both in neural networks and statistical signal processing, because one can often obtain excellent results using them provided that the assumed model is of correct type. They allow utilization of the available prior information by modeling them using suitable prior distributions, and a fully Bayesian treatment makes it possible to select an optimal model order, making such methods robust against overfitting. The main disadvantages of fully Bayesian estimation methods have been their often quite high computational load and intractable computations without approximations. These obstacles have prevented their application to realistic unsupervised or blind learning problems where the number of unknown parameters to be estimated grows easily large.

Variational Bayesian learning, also called Bayesian ensemble learning [58], utilizes an approximation which is fitted to the posterior distribution of the parameter(s) to be estimated. The approximative distribution is often chosen to be Gaussian because of its simplicity and computational efficiency. The mean of this Gaussian distribution provides a point estimate for the unknown parameter considered, and its variance gives a somewhat crude but useful measure of the reliability of the point estimate. The approximative posterior distribution is fitted to the posterior distribution estimated from the data using the Kullback-Leibler information (divergence) [1], [4]. This measures the difference between two probability densities, and is sensitive to the mass of the distributions rather than to some peak value, resulting in robust estimates.

Variational Bayesian methods were first applied to standard linear ICA and BSS in [59], [60], and several research groups have since then used Bayesian approaches to handle various blind problems for linear models; see [16], [1], [18], [61] and the references therein. Valpola (earlier Lappalainen) and his co-authors have introduced several methods based on Bayesian ensemble learning for blind estimation and separation in nonlinear mixture (data) models. In these methods, the nonlinear mapping $\mathbf{f}$ in (2) is modeled using a multilayer perceptron (MLP) network [29] with one nonlinear hidden layer, and the data model (2) contains also additive noise. The necessary regularization for nonlinear BSS is achieved by choosing the model and sources that have most probably generated the observed data.

Assuming that the source signals $\mathbf{s}$ at the input layer of the MLP network have simple Gaussian distributions, one obtains a nonlinear principal component analysis (PCA) solution called nonlinear factor analysis (NFA) [62], [63], [64], [1]. The NFA solution can usually model quite well the nonlinear mixtures (observed data), but it does not yet provide estimates of the independent source signals, because the sources have plain Gaussian distributions in the NFA method. The simplest way to achieve nonlinear BSS is to apply standard linear ICA to the found NFA solution. The quality of this nonlinear BSS solution can be improved still somewhat by continuing Bayesian ensemble learning, but using now a more sophisticed a mixture-of-Gaussians model for the sources. It is well known that suitable mixtures of Gaussian distributions are able to model with sufficient accuracy any source distributions. This method is called Nonlinear Independent Factor Analysis (NIFA).

The NFA and NIFA methods were first introduced in [62], and a more principled theoretical derivation has been presented in [63]. Experimental results with artificially generated data, showing that the NFA method followed by linear ICA and the NIFA method are able to approximate pretty well the true sources, have been presented in [1], [62], [64]. These methods have been applied also to real-world data sets, including 30-dimensional pulp data [1], [62], [64] and speech data [65], but interpretation of the results is somewhat difficult, requiring problem-specific expertise.

Somewhat later on, the NFA method was extended to include a nonlinear dynamic model for the sources in [66]. The developed NDFA (Nonlinear Dynamic Factor Analysis) method is presented thoroughly in [61], and the results obtained thus far have been summarized in [65]. The MATLAB codes for the NFA and NDFA methods are available at the www site [67].

More specifically, the data model used in the NDFA method is

$$\mathbf{x}(t) = \mathbf{f}(\mathbf{s}(t)) + \mathbf{n}(t) \tag{34}$$

$$\mathbf{s}(t) = \mathbf{g}(\mathbf{s}(t-1)) + \mathbf{m}(t) \tag{35}$$

In the latter equation (35), $\mathbf{g}(t)$ is another unknown nonlinear function which controls the dynamics of the sources $\mathbf{s}(t)$, and $\mathbf{m}(t)$ is a similar additive noise term as $\mathbf{n}(t)$ in the static nonlinear data model (34). Similarly as in the NFA method, the function $\mathbf{g}(t)$ is modeled by an MLP network, and the unknown mappings $\mathbf{f}(t)$ and $\mathbf{g}(t)$ as well as the sources are learned using Bayesian ensemble learning. The model (34)–(35) is discussed in detail in [61].

Many real-world data sets can be appropriately described as nonlinear dynamic systems such as (34)–(35), and therefore nonlinear BSS for dynamical systems may in fact have more practical applications than static nonlinear BSS. The first paper about nonlinear dynamical ICA (to our knowledge) is [68], where the authors have considered state-space models and a hyper radial-basis function (RBF) network [29] for nonlinear mixtures. However, the method introduced in [68] is not completely blind, because it partly resorts to supervised learning.

In experiments with difficult chaotic data [66], [61], the NDFA method performed excellently, outperforming for example the prediction results given by nonlinear autoregressive modeling learned by standard back-propagation (see [29], Chapter 15) by an order of magnitude. The NDFA method has been applied also to BSS of biomedical MEG data in [69], and it provided clearly better results than standard linear ICA. In this application, the static data model (34) was a standard linear ICA model, but the dynamic model (35) was nonlinear. The NDFA method has been used not only to blind estimation of the dynamic system and its source signals, but also to detection of changes in the states (sources) of the process in [70], [65]. The method performed again much better than the compared state-of-the-art techniques of change detection.

A problem in particular in the NDFA method but also in the NFA and NIFA methods is that their computational load is still high in problems of realistic size in spite of the efficient Gaussian approximation. Another problem is that the Bayesian ensemble learning procedure may get stuck to a local minimum and requires careful initialization. To combat these problems, a simpler block approach which neglects all posterior dependencies has been recently developed in [71]. The block approach allows straightforward construction and Bayesian ensemble learning of a variety of models, and it is computationally clearly more efficient and robust against local minima. It can be used for learning variance sources for linear and nonlinear models [71], [72], and in [73] we have tested it in nonlinear BSS. The results for artificial data are slightly worse than when using the NFA method followed by linear ICA or the NIFA method, but preliminary experiments with real-world speech data are quite encouraging.

Occasionally, the approximation used in the block method which neglects all posterior dependencies may be too simple for providing the true ICA or BSS solution, leading to inferior performance. This problem and solutions to it are discussed in [74]. Ensemble learning can be accelerated also by applying an improved updating scheme for the parameters to be estimated [75].

B. Other approaches

In this section, we briefly review other methods proposed for nonlinear ICA or BSS.

Already in 1987, Jutten [76] used soft nonlinear mixtures for assessing the robustness and performance of the seminal Hérault-Jutten algorithm. However, Burel [77] was probably the first to introduce an algorithm specifically for nonlinear ICA. His method, restricted to known nonlinearities with unknown parameters and based on back-propagation type neural learning, suffers from huge computational complexity and problems with local minima.

A few years after this, Deco and Parra with their co-authors developed in a series of papers [78], [79], [80], [81], [82] nonlinear ICA methods based on volume conserving symplectic transformations. In fact, the constraint of volume conservation is somewhat arbitrary, and hence these methods are usually not able to recover the original sources.

One of the earliest ideas for achieving general nonlinear BSS (or ICA), introduced first by Pajunen in [83] and then together with co-authors in [84], is to use the well-known self-organizing map (SOM) (see for example [29]) to that end. SOM learns in an unsupervised manner a nonlinear mapping from the data to a usually 2-D rectangular grid. With suitable modifications [83], [84], the mapping provided by SOM is roughly uniformly distributed on the grid. The marginal densities along the sides of the rectangular grid become then statistically independent. The SOM mapping also provides a regularization mechanism needed in nonlinear BSS, because it tries to preserve the structure of the data by using a nonlinear mapping as simple as possible [29], [1].

The SOM-based nonlinear BSS method has been successfully applied to denoising of images corrupted by multiplicative noise in a recent journal paper [85]. A comparison between standard SOM and its modified version which is more suitable for dealing with multiplicative noise is presented together with experimental separation results on test and real images.

In general, the SOM-based nonlinear BSS method is able to approximately separate the sources if their distributions are close to the uniform one, but the results become the poorer the farther away the distributions of the sources are from the uniform one [1]. Another difficulty in using SOM for nonlinear BSS or ICA is that computational complexity increases very rapidly with the number of the sources, limiting the potential application of this method to small-scale problems [1], [85]. Some further results on the applicability of the SOM-based method to linear and nonlinear BSS in simple cases have been given also in [86]. Lin, Grier, and Cowan [87] have independently proposed using SOM for nonlinear ICA and BSS in a different manner by treating ICA as a local computational geometry problem.

The restriction of uniform distributions can be alleviated by using instead of SOM so-called generative topographic mapping (GTM) method, which was introduced in [88] as a principled and theoretically well founded alternative to the somewhat heuristic SOM method. A nonlinear BSS method relying on a slightly modified version of GTM was introduced in [89], and it is discussed somewhat more thoroughly in Section 17.4 of the book [1]. The method requires knowledge of the distributions of the sources which no longer need to be close to the uniform one, but the problem with the curse of the dimensionality remains.

In addition to the variational Bayesian approaches of the previous subsection, MLP networks have been employed in several other nonlinear BSS or ICA methods as flexible models for the nonlinear mixing mapping (2). Autoassociative MLP networks [29] in which the desired output vector of the network is the same as the input mixture vector $\mathbf{x}(t)$ have been tried for this task. Both the generative model (2) and its inversion (3) are learned simultaneously, but separately without utilizing the fact that the models are connected. Autoassociative MLPs have shown some success in nonlinear data representation [29], but generally they suffer from slow learning prone to local minima.

Most works on autoassociative MLPs use point estimates for weights and sources obtained by minimizing the mean-square representation error for the data. It is then impossible to reliably choose the structure of the model, and problems with over- or underfitting can be severe. Hecht-Nielsen [90], [91] proposed so-called replicator networks for universal optimal nonlinear coding of input data. Replicator networks are autoassociate MLP networks, where the data vectors are mapped onto a unit hypercube so that the mapped data is uniformly distributed inside the hypercube. The coordinates of the mapped data on the axes of the hypercube, called natural coordinates, form then in fact a nonlinear ICA solution, even though this has not been noticed in the original papers [90], [91].

Hochreiter and Schmidhuber [92] have used in context with MLP networks a method based on minimum description length, called LOCOCODE. This method does estimate the distribution of the weights, but it has no model for the sources. It is then impossible to measure the description length of the sources. Anyway, the experimental results yielded by the LOCOCODE method show interesting connections with ICA; sometimes the method provides a nonlinear ICA solution, sometimes it does not [92].

Another well-known information theoretic criterion, mutual information, is applied to measuring statistical independence in [30], [93]. In these papers, various MLP network based methods have also been introduced for nonlinear blind separation. In particular, Yang, Amari, and Cichocki [93] deal with extensions of the basic natural gradient method for nonlinear BSS, and furthermore present another extension based on entropy maximization and experiments with post-nonlinear mixtures. This technique has been generalized to mixtures of sigmoidal nonlinearities, allowing an improved fitting to complicated nonlinear mixing functions, in [94]. Yet another paper suggesting a maximum entropy method for nonlinear ICA using a MLP network structure is [95].

Kernel-based methods, in particular kernel ICA [96] and kernel PCA which is a nonlinear extension of standard PCA (see again [29] for a brief description) can also be used as a starting point for developing algorithms for nonlinear ICA and BSS. A first paper on this line of research is [97], where kernel canonical correlation analysis has been suggested for nonlinear ICA and some other extensions of standard linear ICA. A more efficient algorithm with a successful example of blind separation of nonlinearly mixed speech signals has been introduced in [98]. An open problem with these somewhat heuristic kernel methods is how to choose the nonlinear transformation into a higher dimensional space so that the mixtures become roughly linearly separable there.

Tan and Zurada [99] have proposed a radial basis function (RBF) neural network structure for approximating the separating mapping (3). Their contrast function consists of the mutual information as well as of partial moments of the estimated separated sources, which are used to provide the regularization needed in nonlinear BSS. Simulation results are presented for several artificially generated nonlinear mixture sets, confirming the validity of the method introduced in [99].

Marques and Almeida [100] have introduced a pattern repulsion method based on the maximum entropy principle for solving the nonlinear ICA or BSS problems. Their model, whose origins lie in statistical physics, is studied theoretically in [101]. Xiong and Huang [102] propose a generalization of the classic Bell-Sejnowski algorithm which uses power series of the nonlinear mixtures to approximate the Taylor expansion of the separating mapping (3). Genetic algorithms have been considered for improving the estimation of parameters of the separating mapping in [103].

A technique inspired by properties of electric fields, applicable for nonlinear ICA and some other extensions of ICA, has been suggested in [104]. It can be viewed also as a method for constructing local density approximations to the joint and factorial distributions, providing a more rigorous theoretical foundation to this method [105].

Several researchers have considered the possibility of applying their methods to nonlinear ICA or BSS without presenting a complete study including experimental results. Xu [106] has developed a general Bayesian Ying-Yang framework which could be applied to nonlinear ICA. Hinton et al. [107] have interpreted ICA in a novel way as an energy-based probability density model, mentioning that it is easy to extend the approach to nonlinear ICA models. A related paper is [108]. Chen and Gopinath propose an iterative Gaussianization technique in [109] which could according to them provide a computationally efficient solution to nonlinear ICA. The idea of applying Gaussianization methods for separating post-nonlinear mixtures has been introduced independently in [50], [49], see Section III.B.

C. Local ICA and BSS

Nonlinear independent component analysis or blind source separation are generally both computationally and conceptually difficult problems. Therefore, local linear ICA and/or BSS methods have received some attention recently as a practical compromise between linear ICA and completely nonlinear ICA or BSS. These methods are more general than standard linear ICA in that several different linear ICA models are used to describe the observed data. The local linear ICA models can be either overlapping, as in the promising Bayesian mixture-of-ICA methods introduced in [110], [111], or nonoverlapping, as in the clustering-based methods proposed in [112], [113]. In [114], [115], variational Bayesian approach has been used for determining the number of local independent components in high-dimensional data sets, with a successful application to a difficult real-world medical data. Feature detection tools from image analysis have been used for estimating local ICA coordinate systems in [116]. In [117], Lin presents interesting theoretical considerations based on local geometric structure which can be used for BSS of nonlinear mixtures.

V. Concluding remarks

In this paper, we have considered ICA and BSS problems for nonlinear data models. In that case, ICA is characterized by huge indeterminacies, and extra constraints or regularization are necessary for actually achieving solutions which coincide to source separation.

Two main results can be stated. 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 (II-G.2). Second, prior information on sources, for example bounded or temporally correlated sources, can simplify the algorithms or reduce the indeterminacies in the solutions.

Another promising method consists in regularizing the solution using a fully Bayesian variational ensemble learning approach. It tries to find the sources and the mapping that have most probably generated the observed data. The ensemble learning method allows nonlinear source separation for problems of realistic size, and it can be easily extended in various directions.

A lot of work remains to be done in studying the nonlinear ICA and BSS problems. First, further studies are needed on the separability problem. Second, these nonlinear problems are difficult and ill-posed without a suitable regularization, and we feel that all the available information should be used whenever possible. For example, incorporation of temporal statistics can be quite helpful. Moreover, 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.

Acknowledgments. This work has been supported by the European Commission project BLISS (IST-1999-14190) and the Academy of Finland. The authors thank their colleagues, in particular Drs. A. Taleb and H. Valpola, for contributing many of the results presented in this paper and for useful co-operation and comments.

References

[1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. J. Wiley, 2001.

[2] P. Comon, “Independent component analysis, a new concept?” *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[3] J.-F. Cardoso, “Blind signal separation: statistical principles,” *Proc. of the IEEE*, vol. 9, no. 10, pp. 2009–2025, 1998.

[4] A. Cichocki and S.-I. Amari, *Adaptive Blind Signal and Image Processing - Learning Algorithms and Applications*. J. Wiley, 2002.

[5] C. Jutten and A. Taleb, “Source separation: from dusk till dawn,” in *Proc. 2nd Int. Workshop on Independent Component Analysis and Blind Source Separation (ICA2000)*, Helsinki, Finland, 2000, pp. 15–26.

[6] C. Jutten and J. Hérault, “Blind separation of sources, Part I: an adaptive algorithm based on a neuromimetic architecture,” *Signal Processing*, vol. 24, no. 1, pp. 1–10, 1991.

[7] J.-F. Cardoso and A. Souloumiac, “Blind beamforming for non gaussian signals,” *IEE Proceedings-F*, vol. 140, no. 6, pp. 362–370, 1993.

[8] A. Cichocki, R. Unbehauen, and E. Rummert, “Robust learning algorithm for blind separation of signals,” *Electronics Letters*, vol. 30, no. 17, pp. 1386–1387, 1994.

[9] A. Bell and T. Sejnowski, “An information-maximization approach to blind separation and blind deconvolution,” *Neural Computation*, vol. 7, no. 6, 1995.

[10] J.-F. Cardoso and B. Laheld, “Equivariant adaptive source separation,” *IEEE Trans. on Signal Processing*, vol. 44, no. 12, pp. 3017–3030, 1996.

[11] S. Amari, A. Cichocki, and H. Yang, “A new learning algorithm for blind signal separation,” in *Advances in Neural Information Processing Systems*, Denver (Colorado), December 1996, pp. 757–763.

[12] A. Hyvärinen and E. Oja, “A fast fixed-point algorithm for independent component analysis,” *Neural Computation*, vol. 9, no. 7, pp. 1483–1492, 1997.

[13] J.-F. Cardoso, C. Jutten, and P. Loubaton, Eds., *Proc. of the First Int. Workshop on Independent Component Analysis and Signal Separation*, Aussois, France, January 1999.

[14] P. Pajunen and J. Karhunen, Eds., *Proc. of the 2nd Int. Workshop on Independent Component Analysis and Blind Signal Separation*, Helsinki, Finland, June 2000.

[15] T.-W. Lee, S. Makeig, and T. Sejnowski, Eds., *Proc. of the 3rd Int. Conf. on Independent Component Analysis and Signal Separation*, San Diego, CA, USA, December 2001.

[16] M. Girolami, Ed., *Advances in Independent Component Analysis*. Springer-Verlag, 2000.

[17] S. Haykin, Ed., *Unsupervised Adaptive Filtering, Vol. I: Blind Source Separation*. Wiley, 2000.

[18] S. Roberts and R. Everson, Eds., *Independent Component Analysis: Principles and Practice*. Cambridge Univ. Press, 2001.

[19] L. Tong, V. Soon, Y. Huang, and R. Liu, “AMUSE: a new blind identification algorithm,” in *Proc. of IEEE Int. Symp. on Circuits and Systems (ISCAS'90)*, New Orleans, LA, USA, 1990.

[20] A. Belouchrani, K. A. Meraim, J.-F. Cardoso, and E. Moulines, “A blind source separation technique based on second order statistics,” *IEEE Trans. on Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.

[21] K. Matsuoka, M. Ohya, and M. Kawamoto, “A neural net for blind separation of nonstationary signals,” *Neural Networks*, vol. 8, no. 3, pp. 411–419, 1995.

[22] D. T. Pham and J.-F. Cardoso, “Blind separation of instantaneous mixtures of nonstationary sources,” *IEEE Trans. on Signal Processing*, vol. 49, no. 9, pp. 1837–1848, 2001.

[23] A. Hyvärinen and P. Pajunen, “Nonlinear independent component analysis: Existence and uniqueness results,” *Neural Networks*, vol. 12, no. 3, pp. 429–439, 1999.

[24] A. Taleb and C. Jutten, “Source separation in post-nonlinear mixtures,” *IEEE Trans. on Signal Processing*, vol. 47, no. 10, pp. 2807–2820, 1999.

[25] A. Taleb, “A generic framework for blind source separation in structured nonlinear models,” *IEEE Trans. on Signal Processing*, vol. 50, no. 8, pp. 1819–1830, 2002.

[26] J. Eriksson and V. Koivunen, “Blind identifiability of class of nonlinear instantaneous ICA models,” in *Proc. of the XI European Signal Proc. Conf. (EUSIPCO 2002)*, vol. 2, Toulouse, France, September 2002, pp. 7–10.

[27] G. Darmois, “Analyse des liaisons de probabilité,” in *Proc. of Int. Statistics Conferences 1947*, vol. III A, Washington (D.C.), 1951, p. 231.

[28] E. Lukacs, “A characterization of the Gamma distribution,” *Annals of Mathematical Statistics*, no. 26, pp. 319–324, 1955.

[29] S. Haykin, *Neural Networks - A Comprehensive Foundation*, 2nd ed. Prentice Hall, 1998.

[30] L. Almeida, "Linear and nonlinear ICA based on mutual information," in *Proc. IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (AS-SPCC)*, Lake Louise, Canada, October 2000, pp. 117–122.

[31] M. Babaie-Zadeh, "On blind source separation in convolutive and nonlinear mixtures," Ph.D. dissertation, INPG, Grenoble, France, September 2002.

[32] A. Parashiv-Ionescu, C. Jutten, A. Ionescu, A. Chovet, and A. Rusu, "High performance magnetic field smart sensor arrays with source separation," in *Proc. of the 1st Int. Conf. on Modeling and Simulation of Microsystems, MSM 98*, Santa Clara, CA, USA, April 1998, pp. 666–671.

[33] S. Prakriya and D. Hatzinakos, "Blind identification of LTI-ZMNL-LTI nonlinear channel models," *IEEE Trans. on Signal Processing*, vol. 43, no. 12, pp. 3007–3013, December 1995.

[34] M. Korenberg and I. Hunter, "The identification of nonlinear biological systems: LNL cascade models," *Biol. Cybernetics*, vol. 43, no. 12, pp. 125–134, December 1995.

[35] M. Babaie-Zadeh, C. Jutten, and K. Nayebi, "Separating convolutive post non-linear mixtures," in *Proc. of the 3rd Workshop on Independent Component Analysis and Signal Separation (ICA2001)*, San Diego (California, USA), 2001, pp. 138–143.

[36] A. Taleb, J. Sole, and C. Jutten, "Quasi-nonparametric blind inversion of Wiener systems," *IEEE Trans. on Signal Processing*, vol. 49, no. 5, pp. 917–924, 2001.

[37] F. Theis and E. Lang, "Maximum entropy and minimal mutual information in a nonlinear model," in *Proc. Int. Conf. on Independent Component Analysis and Signal Separation (ICA2001)*, San Diego, CA, USA, 2001, pp. 669–674.

[38] L. Almeida, C. Jutten, and H. Valpola, "Realistic models of nonlinear mixtures," HUT, INESC, INPG, BLISS (IST1999-14190) Project report D5, May 2001.

[39] R. Popovic, *Hall-Effect Devices*. Bristol: Adam Hilger, 1991.

[40] A. Kagan, Y. Linnik, and C. Rao, "Extension of Darmois-Skitovic theorem to functions of random variables satisfying an addition theorem," *Communications in Statistics*, vol. 1, no. 5, pp. 471–474, 1973.

[41] J. Aczel, *Lectures on Functional Equations and Their Applications*. New-York: Academic Press, 1966.

[42] M. Babaie-Zadeh, C. Jutten, and K. Nayebi, "A geometric approach for separating post nonlinear mixtures," in *Proc. of the XI European Signal Processing Conf. (EUSIPCO 2002)*, vol. II, Toulouse, France, 2002, pp. 11–14.

[43] S. Hosseini and C. Jutten, "On the separability of nonlinear mixtures of temporally correlated sources," *IEEE Signal Processing Letters*, 2003, to appear.

[44] A. Taleb and C. Jutten, "Nonlinear source separation: The postlinear mixtures," in *Proc. Europ. Symp. on Artificial Neural Networks (ESANN'97)*, Bruges, Belgium, 1997, pp. 279–284.

[45] ——, "Batch algorithm for source separation in post-nonlinear mixtures," in *Proc. First Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, Aussois, France, 1999, pp. 155–160.

[46] T.-W. Lee, B. Koehler, and R. Orglmeister, "Blind source separation of nonlinear mixing models," in *Neural Networks for Signal Processing VII*. IEEE Press, 1997, pp. 406–415.

[47] S. Achard, D. T. Pham, and C. Jutten, "Blind source separation in post nonlinear mixtures," in *Proceedings of ICA 2001*, San Diego, CA, USA, 2001, pp. 295–300.

[48] A. Ziehe, M. Kawanabe, S. Harmeling, and K.-R. Müller, "Separation of post-nonlinear mixtures using ACE and temporal decorrelation," in *Proc. Int. Conf. on Independent Component Analysis and Signal Separation (ICA2001)*, San Diego, USA, 2001, pp. 433–438.

[49] J. Solé, M. Babaie-Zadeh, C. Jutten, and D.-T. Pham, "Improving algorithm speed in PNL mixture separation and Wiener system inversion," Nara (Japan), April 2003, submitted to the ICA2003 Symposium.

[50] A. Ziehe, M. Kawanabe, S. Harmeling, and K.-R. Müller, "Blind separation of post-nonlinear mixtures using gaussianizing transformations and temporal decorrelation," Nara (Japan), April 2003, submitted to the ICA2003 Symposium.

[51] H. Peng, Z. Chi, and W. Siu, "A semi-parametric hybrid neural model for nonlinear blind signal separation," *Int. J. of Neural Systems*, vol. 10, no. 2, pp. 79–94, 2000.

[52] M. Solazzi, R. Parisi, and A. Uncini, "Blind source separation in nonlinear mixtures by adaptive spline neural networks," in *Proc. Int. Conf. on Independent Component Analysis and Signal Separation (ICA2001)*, San Diego, CA, USA, 2001, pp. 254–259.

[53] C. Puntonet, M. Alvarez, A. Prieto, and B. Prieto, "Separation of sources in a class of post-nonlinear mixtures," in *Proc. 6th European Symp. on Artificial Neural Networks (ESANN'98)*, Bruges, Belgium, April 1998, pp. 321–326.

[54] J. Sole, C. Jutten, and A. Taleb, "Parametric approach to blind deconvolution of nonlinear channels," *Neurocomputing*, vol. 48, pp. 339–355, 2002.

[55] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley Series in Telecommunications, 1991.

[56] S. Sénécal and P.-O. Amblard, "Blind equalization of a nonlinear satellite system using MCMC simulation methods," *Applied Signal Processing*, vol. 9, pp. 46–57, 2002.

[57] M. Babaie-Zadeh, C. Jutten, and K. Nayebi, "A minimization-projection approach for separating post nonlinear mixtures," Nara, Japan, April 2003, submitted to the 4th Int. Symp. on Independent Component Analysis and Blind Signal Separation (ICA 2003).

[58] H. Lappalainen and J. Miskin, "Ensemble learning," in *Advances in Independent Component Analysis*, M. Girolami, Ed. Berlin: Springer-Verlag, 2000, pp. 75–92.

[59] H. Attias, "Independent factor analysis," *Neural Computation*, vol. 11, no. 4, pp. 803–851, 1999.

[60] H. Lappalainen, "Ensemble learning for independent component analysis," in *Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, Aussois, France, 1999, pp. 7–12.

[61] H. Valpola and J. Karhunen, "An unsupervised ensemble learning method for nonlinear dynamic state-space models," *Neural Computation*, vol. 14, no. 11, pp. 2647–2692, 2002.

[62] H. Lappalainen and A. Honkela, "Bayesian nonlinear independent component analysis by multi-layer perceptrons," in *Advances in Independent Component Analysis*, M. Girolami, Ed. Springer-Verlag, 2000, pp. 93–121.

[63] H. Valpola, "Nonlinear independent ccmponent analysis using ensemble learning: Theory," in *Proc. 2nd Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000)*, Helsinki, Finland, 2000, pp. 251–256.

[64] H. Valpola, X. Giannakopoulos, A. Honkela, and J. Karhunen, "Nonlinear independent component analysis using ensemble learning: Experiments and discussion," in *Proc. 2nd Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000)*, Helsinki, Finland, 2000, pp. 351–356.

[65] H. Valpola, E. Oja, A. Ilin, A. Honkela, and J. Karhunen, "Nonlinear blind source separation by variational Bayesian learning," *IEICE Transactions (Japan)*, vol. E86-A, no. 3, March 2003, to appear in a Special Section on Blind Signal Processing (Invited paper).

[66] H. Valpola, A. Honkela, and J. Karhunen, "An ensemble learning approach to nonlinear dynamic blind source separation using state-space models," in *Proc. Int. Joint Conf. on Neural Networks (IJCNN'02)*, Honolulu, HI, USA, 2002, pp. 460–465.

[67] H. Valpola, A. Honkela, and X. Giannakopoulos, "MATLAB codes for the NFA and NDFA algorithms," *http://www.cis.hut.fi/projects/ica/bayes/*, 2002.

[68] A. Cichocki, L. Zhang, S. Choi, and S.-I. Amari, "Nonlinear dynamic independent component analysis using state-space and neural network models," in *Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, Aussois, France, 1999, pp. 99–104.

[69] J. Särelä, H. Valpola, and R. V. E. Oja, "Dynamical factor analysis of rhythmic magnetoencephalographic activity," in *Proc. Int. Conf. on Independent Component Analysis and Signal Separation (ICA2001)*, San Diego, CA, USA, 2001, pp. 451–456.

[70] A. Ilin, H. Valpola, and E. Oja, "Detecting process state changes by nonlinear blind source separation," in *Proc. Int. Conf. on Independent Component Analysis and Signal Separation (ICA2001)*, San Diego, USA, December 2001, pp. 704–709.

[71] H. Valpola, T. Raiko, and J. Karhunen, "Building blocks for hierarchical latent variable models," in *Proc. 3rd Int. Conf. on Independent Component Analysis and Signal Separation (ICA2001)*, San Diego, USA, 2001, pp. 710–715.

[72] H. Valpola, M. Harva, and J. Karhunen, "Hierarchical models of variance sources," in *Proc. of 4th Int. Symp. on Independent Component Analysis and Blind Source Separation (ICA2003)*, A. Cichocki and N. Murata, Eds., Nara, Japan, April 2003, submitted.

[73] H. Valpola, T. Östman, and J. Karhunen, "Nonlinear independent factor analysis by hierarchical models," in *Proc. of 4th Int. Symp. on Independent Component Analysis and Blind Source Separation (ICA2003)*, A. Cichocki and N. Murata, Eds., Nara, Japan, April 2003, submitted.

[74] A. Ilin and H. Valpola, "On the effect of the form of the posterior approximation in variational learning of ICA models," in *Proc. 4th Int. Symp. on Independent Component Analysis and Blind Signal Separation (ICA2003)*, Nara, Japan, April 2003, submitted.

[75] A. Honkela, "Speeding up cyclic update schemes by pattern searches," in *Proc. of the 9th Int. Conf. on Neural Inform. Processing (ICONIP'02)*, Singapore, 2002, pp. 512–516.

[76] C. Jutten, "Calcul neuromimétique et traitement du signal, analyse en composantes indépendantes," Ph.D. dissertation, INPG, Univ. Grenoble, France, 1987, (in French).

[77] G. Burel, "Blind separation of sources: A nonlinear neural algorithm," *Neural Networks*, vol. 5, no. 6, pp. 937–947, 1992.

[78] G. Deco and W. Brauer, "Nonlinear higher-order statistical decorrelation by volume-conserving neural architectures," *Neural Networks*, vol. 8, pp. 525–535, 1995.

[79] L. Parra, G. Deco, and S. Miesbach, "Redundancy reduction with information-preserving nonlinear maps," *Network*, vol. 6, pp. 61–72, 1995.

[80] L. Parra, "Symplectic nonlinear component analysis," in *Advances in Neural Inform. Proc. Systems*. Cambridge, Mass.: MIT Press, 1996, vol. 8, pp. 437–443.

[81] G. Deco and D. Obradovic, *An Information-Theoretic Approach to Neural Computing*. Springer Verlag, 1996.

[82] L. Parra, G. Deco, and S. Miesbach, "Statistical independence and novelty detection with information-preserving nonlinear maps," *Neural Computation*, vol. 8, pp. 260–269, 1996.

[83] P. Pajunen, "Nonlinear independent component analysis by self-organizing maps," in *Proc. of the 1996 Int. Conf. on Artificial Neural Networks (ICANN'96)*, C. von der Malsburg et al., Ed. Bochum, Germany: Springer-Verlag, 1996, pp. 815–820.

[84] P. Pajunen, A. Hyvärinen, and J. Karhunen, "Nonlinear blind source separation by self-organizing maps," in *Proc. Int. Conf. on Neural Information Processing*, Hong Kong, 1996, pp. 1207–1210.

[85] M. Haritopoulos, H. Yin, and N. Allison, "Image denoising using self-organizing map-based nonlinear independent component analysis," *Neural Networks*, vol. 15, pp. 1085–1098, 2002.

[86] M. Herrmann and H. Yang, "Perspectives and limitations of self-organizing maps," in *Proc. Int. Conf. on Neural Information Processing (ICONIP'96)*, Hong Kong, 1996, pp. 1211–1216.

[87] J. Lin, D. Grier, and J. Cowan, "Faithful representation of separable distributions," *Neural Computation*, vol. 9, no. 6, pp. 1305–1320, 1997.

[88] C. Bishop, M. Svensen, and C. Williams, "GTM: The generative topographic mapping," *Neural Computation*, vol. 10, pp. 215–234, 1998.

[89] P. Pajunen and J. Karhunen, "A maximum likelihood approach to nonlinear blind source separation," in *Proc. of the 1997 Int. Conf. on Artificial Neural Networks (ICANN'97)*, Lausanne, Switzerland, 1997, pp. 541–546.

[90] R. Hecht-Nielsen, "Replicator neural networks for universal optimal source coding," *Science*, vol. 269, pp. 1860–1863, 1995.

[91] ——, "Data manifolds, natural coordinates, replicator neural networks, and optimal source coding," in *Proc. Int. Conf. on Neural Information Processing (ICONIP'96)*, Hong Kong, 1996, pp. 41–45.

[92] S. Hochreiter and J. Schmidhuber, "Feature extraction through LOCOCODE," *Neural Computation*, vol. 11, no. 3, pp. 679–714, 1999.

[93] H. Yang, S.-I. Amari, and A. Cichocki, "Information-theoretic approach to blind separation of sources in non-linear mixture," *Signal Processing*, vol. 64, no. 3, pp. 291–300, 1998.

[94] A. Koutras, E. Dermatas, and G. Kokkinakis, "Neural network based blind source separation of non-linear mixtures," in *Proc. of the 2001 Int. Conf. on Artificial Neural Networks (ICANN'01)*. Wien, Austria: Springer-Verlag, 2001, pp. 561–567.

[95] J. Fisher and J. Principe, "Entropy manipulation of arbitrary nonlinear mappings," in *Neural Networks for Signal Processing VII*, J. P. et al. Ed. New York: IEEE Press, 1997, pp. 14–23.

[96] F. Bach and M. Jordan, "Kernel independent component analysis," *Journal of LMachine Learning Research*, vol. 3, pp. 1–48, 2002.

[97] C. Fyfe and P. Lai, "ICA using kernel canonical correlation analysis," in *Proc. 2nd Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000)*, Helsinki, Finland, 2000, pp. 279–284.

[98] S. Harmeling, A. Ziehe, B. Blankertz, and K.-R. Müller, "Nonlinear blind source separation using kernel feature spaces," in *Proc. Int. Conf. on Independent Component Analysis and Signal Separation (ICA2001)*, San Diego, USA, 2001, pp. 102–107.

[99] Y. Tan, J. Wang, and J. Zurada, "Nonlinear blind source separation using a radial basis function network," *IEEE Trans. on Neural Networks*, vol. 12, no. 1, pp. 124–134, January 2001.

[100] G. Marques and L. Almeida, "Separation of nonlinear mixtures using pattern repulsion," in *Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, Aussois, France, 1999, pp. 277–282.

[101] F. Theis, C. Bauer, C. Puntonet, and E. Lang, "Pattern repulsion revisited," in *Proc. Int. Work-Conference on Artificial and Natural Neural Networks (IWANN2001)*, J. Mira and A. Prieto, Eds. Granada, Spain: Springer-Verlag, June 2001, pp. 778–785.

[102] Z. Ziong and T. Huang, "Nonlinear independent component analysis (ICA) using power series and application to blind source separation," in *Proc. 3rd Int. Conf. on Independent Component Analysis and Signal Separation (ICA2001)*, San Diego, CA, USA, 2001, pp. 680–685.

[103] F. Rojas, I. Rojas, R. Clemente, and C. Puntonet, "Nonlinear blind source separation using genetic algorithms," in *Proc. Int. Conf. on Independent Component Analysis and Signal Separation (ICA2001)*, San Diego, USA, 2001, pp. 400–405.

[104] S. Hochreiter and M. Mozer, "An electric field approach to independent component analysis," in *Proc. Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000)*, Helsinki, Finland, 2000, pp. 45–50.

[105] ——, "Beyond maximum likelihood and density estimation: A sample-based criterion for unsupervised learning of complex models," in *Advances in Neural Information Processing Systems*, T. Leen, T. Dietterich, and V. Tresp, Eds., vol. 13. The MIT Press, 2001, pp. 535–541.

[106] L. Xu, "Temporal BYY learning for state space approach, hidden markov model, and blind source separation," *IEEE Trans. on Signal Processing*, vol. 48, no. 7, pp. 2132–2144, 2000.

[107] G. Hinton, M. Welling, Y. Teh, and S. Osindero, "A new view of ICA," in *Proc. Int. Conf. on Independent Component Analysis and Signal Separation (ICA2001)*, San Diego, CA, USA, 2001, pp. 746–751.

[108] T. Marks and J. Movellan, "Diffusion networks, product of experts, and factor analysis," in *Proc. Int. Conf. on Independent Component Analysis and Signal Separation (ICA2001)*, San Diego, CA, USA, 2001, pp. 481–485.

[109] S. Chen and R. Gopinath, "Gaussianization," in *Advances in Neural Inform. Proc. Systems*, T. Leen, T. Dietterich, and V. Tresp, Eds., vol. 13. The MIT Press, 2001, pp. 423–429.

[110] T.-W. Lee, M. Lewicki, and T. Sejnowski, "ICA mixture models for unsupervised classification of non-gaussian sources and automatic context switching in blind signal separation," *IEEE Trans. Pattern Recognition and Machine Intelligence*, vol. 22, no. 10, pp. 1–12, 2000.

[111] T.-W. Lee and M. Lewicki, "Image processing methods using ICA mixture models," in *Independent Component Analysis: Principles and Practice*, S. Roberts and R. Everson, Eds. Cambridge University Press, 2001, pp. 234–253.

[112] J. Karhunen, S. Malaroiou, and M. Ilmoniemi, "Local linear independent component analysis based on clustering," *Int. J. of Neural Systems*, vol. 10, no. 6, pp. 439–451, 2000.

[113] F. Palmieri and A. Budillon, "Multi-class independent component analysis (MUCICA)," in *Advances in Independent Component Analysis*, M. Girolami, Ed. Springer-Verlag, 2000, pp. 145–160.

[114] K. Chan, T.-W. Lee, and T. Sejnowski, "Variational learning of clusters of undercomplete nonsymmetric independent components," in *Proc. Int. Conf. on Independent Component Analysis and Signal Separation (ICA2001)*, San Diego, CA, USA, 2001, pp. 492–497.

[115] ——, "Variational learning of clusters of undercomplete nonsymmetric independent components," *J. of Machine Learning Research*, vol. 3, pp. 99–114, August 2002.

[116] J. Lin, D. Grier, and J. Cowan, "Feature extraction approach to blind source separation," in *Neural Networks for Signal Processing VII*, J. P. et al., Ed. New York: IEEE Press, 1997, pp. 398–405.

[117] J. K. Lin, "Factorizing probability density functions: Generalizing ICA," in *Proc. First Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, Aussois, France, 1999, pp. 313–318.