BLISS

IST-1999-14190

Blind Source Separation and Applications

Technical report on Separation methods for nonlinear mixtures

Deliverable D29

Report Version: Final
Report Preparation Date: October 8, 2003
Report Contributors: HUT, INPG, FhG, INESC
Classification: Public
Contract Start Date: 1 June 2000 Duration: 41 months
Project Co-ordinator: INPG
Partners: HUT, INPG, FhG, INESC, McMaster University

Project funded by the European Community under the "Information Society Technologies" Programme (1998-2002)

Separation methods for nonlinear mixtures

Christian Jutten (INPG), Juha Karhunen (HUT),
Luis Almeida (INESC), Stefan Harmeling (FhG)

October 8, 2003

Abstract

This subproject considered ICA and BSS problems for nonlinear data models. In that case, ICA is characterized by severe indeterminacies, which makes the BSS problem ill-posed. Hence some kind of regularization is necessary for actually achieving solutions which can be interpreted as underlying sources. The problem of *regularization* has been analyzed from different perspectives and resulted in powerful methods. For example, explicit constraints on the model class are used in the case of post-nonlinear mixtures, where the nonlinearity is modeled as a component-wise distortion of linear mixtures. More implicit constraints are implemented in kernel-based methods like kTDSEP, where the choice of the kernel parameters regularizes the complexity of the nonlinearities.

Also, exploiting prior information on the sources, for example bounded or temporally correlated sources, reduces the indeterminacies in the solutions and leads to simplified algorithms. 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 given the prior. The ensemble learning method allows nonlinear source separation for problems of realistic size, and it can be easily extended in various directions.

It has been argued that regularization based on an assumption of smoothness of the mixture could overcome the ill-posedness of nonlinear BSS. Although it is still a matter of debate under what exact conditions this is true, several examples of successful source recovery through the MISEP method, using smoothness for regularization, have been presented.

Last but not least, there has been a high degree of cooperation among the partners to compare the results and to promote combined approaches.

1 Introduction: nonlinear ICA and BSS problems

When linear ICA models fail, a natural extension is to consider nonlinear models. For instantaneous mixtures, the nonlinear mixing model has the general form

$$\mathbf{x} = \mathcal{F}(\mathbf{s}) \tag{1}$$

where $\mathbf{x}$ and $\mathbf{s}$ denote the data and source vectors, respectively, 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 $\mathcal{G} : \mathcal{R}^n \rightarrow \mathcal{R}^n$ that yields components

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

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 non-unique. 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 [39, 38, 44].

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 (1), 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.

In the remainder of this report we present a summary of the achieved results in this subproject. The details can be found in the attached publications which are listed in section 2. In the following sections we give a brief overview of the work on post-nonlinear mixtures (see section 3) and on general nonlinear mixtures (see section 4). Finally, the conclusions summarize the major achievements of this subproject.

2 Publications included into the deliverable

Publication 1 [77] 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, pp. 532-541.
In this publication, a general approach for blind separation of sources from their nonlinear mixtures is presented. Multilayer perceptrons are used as nonlinear generative models for the data, and variational Bayesian (ensemble) learning is applied for finding the sources. The variational Bayesian technique automatically provides a reasonable regularization of the nonlinear blind source separation problem. In this publication, we first consider a static nonlinear mixing model, with a successful application to real-world speech data compression. Then we discuss extraction of sources from nonlinear dynamic processes,

and detection of abrupt changes in the process dynamics. In a difficult test problem with chaotic data, our approach clearly outperforms currently available nonlinear prediction and change detection techniques. The proposed techniques are computationally demanding, but they can be applied to blind nonlinear problems of higher dimensions than other existing approaches.

Publication 2 [78] 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, pp. 257-262.
We construct in this publication a hierarchical nonlinear model for nonlinear factor analysis based on the building blocks introduced earlier by us. The resulting method is called hierarchical nonlinear factor analysis (HNFA). The variational Bayesian learning algorithm used in the HNFA method has a linear computational complexity, and it is able to infer the structure of the model in addition to estimating the unknown parameters. We show how sources can be separated from their nonlinear mixtures by first estimating a nonlinear subspace using the HNFA method, and then rotating the found subspace using standard linear independent component analysis. Experimental results show that the cost function minimized during learning predicts well the quality of the estimated subspace.

Publication 3 [55] T. Raiko, H. Valpola, T. Östman, and J. Karhunen, "Missing values in hierarchical nonlinear factor analysis", in *Proc. of Int. Conf. on Artificial Neural Networks and Neural Information Processing (ICANN/ICONIP 2003)*, Istanbul, Turkey, June 26-29, 2003, pp. 185-188.
In this paper, the properties of the hierarchical nonlinear factor analysis (HNFA) method introduced in Publication 2 are studied further by reconstructing missing values. To compare HNFA with other methods, missing values of speech spectrograms have been reconstructed using HNFA, a nonlinear factor analysis method introduced earlier by us (see Publication 1), linear factor analysis, and the self-organizing map. Experimental results suggest that the capacity of the HNFA method to handle nonlinear problems lies between the nonlinear factor analysis and linear factor analysis methods. The HNFA method provides better reconstructions than linear factor analysis, and is more reliable and computationally much more efficient than our previous nonlinear factor analysis method.

Publication 4 [31, 29] S. Harmeling, A. Ziehe, M. Kawanabe, and K.-R. Müller, "Kernel-based nonlinear blind source separation", *Neural Computation*, no. 15, 2003, pp. 1089-1124, attached is Technical report 1/2002,European Commission Research Project BLISS (Blind Source Separation and Applications), IST-1999-14190, 2003.
In this article, FhG proposes kTDSEP, a kernel-based algorithm for nonlinear blind source separation (BSS). It combines complementary research fields: (1) kernel feature spaces and (2) BSS using temporal information. This yields an efficient algorithm for nonlinear BSS with invertible nonlinearity. Key assumptions are that the kernel feature space is chosen rich enough to approximate the nonlinearity and that signals of interest contain temporal information. The reported experiments demonstrate the excellent performance and efficiency of the kTDSEP algorithm for several problems of nonlinear BSS, also for more than two sources.

Publication 5 [82] A. Ziehe, M. Kawanabe, S. Harmeling, and K.-R. Müller, "Blind Separation of Post-Nonlinear Mixtures Using Gaussianizing Transformations and Temporal Decorrelation", *Proc. Int. Conf. on Independent Component Analysis and Signal Separation (ICA2003)*, pages 269–274, Nara, Japan, 2003.
In this paper, FhG proposes two methods that reduce the post-nonlinear blind source separation problem (PNL BSS) to a linear BSS problem. The first method is based on the alternating conditional expectations (ACE algorithm). The second method is a Gaussianizing transformation, which is motivated by the fact that linearly mixed signals before nonlinear transformation are approximately Gaussian distributed. After equalizing the nonlinearities, temporal decorrelation separation (TDSEP) allows us to recover the source signals. Numerical simulations testing "ACE-TD" and "Gauss-TD" on realistic examples are performed with excellent results.

Publication 6 [40] A. Ilin, A. Honkela, S. Achard, and C. Jutten, "The comparison of the blind source separation methods developed at HUT and INPG", Technical Report 2/2003, European Commission Research Project BLISS (Blind Source Separation and Applications), IST-1999-14190, 2003.
As a collaboration between HUT and INPG, two approaches for solving the nonlinear blind source separation (BSS) problem were compared on PNL mixtures: the Bayesian methods developed at HUT and the BSS methods for post-nonlinear (PNL) mixtures developed at INPG. The report also introduces a new Bayesian algorithm for BSS in PNL mixtures, for which promising results are obtained.

Publication 7 [34] A. Honkela, S. Harmeling, L. Lundqvist, and H. Valpola, "Using kernel PCA for initialization of nonlinear factor analysis", Technical Report 1/2003, European Commission Research Project BLISS (Blind Source Separation and Applications), IST-1999-14190, 2003.
The nonlinear factor analysis (NFA) method by Lappalainen and Honkela (2000) [49] is initialized with linear principal component analysis (PCA). Because of the multilayer perceptron (MLP) network used to model the nonlinearity, the method is susceptible to local minima and therefore sensitive to the initialization used. As the method is used for nonlinear separation, the linear initialization may in some cases lead it astray. As a collaboration between HUT and FhG FIRST, we studied how kernel PCA (KPCA) can be used to initialize NFA. Experiments show that this combination can produce significantly better initializations than linear PCA.

Publication 8 [6] L. Almeida, "Faster training in nonlinear ICA using MISEP", in *Proc. Int. Worksh. Independent Component Analysis and Blind Signal Separation*, pages 113–118, Nara, Japan, 2003.
MISEP has been proposed as a generalization of the INFOMAX method in two directions: (1) handling of nonlinear mixtures, and (2) learning the nonlinearities to be used at the outputs, making the method suitable for handling components with a wide range of statistical distributions. This approach is special in that it can be used for general nonlinearities and in that it employs neural networks as the main architecture. It has been shown to be able to handle moderately nonlinear mixtures of up to 10 sources, and to be able to perform nonlinear BSS by using smoothness for regularization.

Publication 9 [1] S. Achard, D.T. Pham, C. Jutten, " Blind source Separation in Post

Nonlinear Mixtures", *Proceeding of ICA 2001 Conference*, San Diego (California, USA), December 2001, pp. 295-300.
In this publication, we implement alternative algorithms to that of Taleb and Jutten for blind source separation in post non-linear mixtures. We use the same mutual information criterion, but we exploit its invariance with respect to translation for deriving its relative gradient in terms of the derivatives of the nonlinear transforms. Then, we design an algorithm, based on a piece-wise parameterization of the nonlinear functions. The algorithm requires estimation of the score functions. In this purpose, we propose a new method for score function estimation.

Publication 10 [**12**] M. Babaie-Zadeh, C. Jutten, K. Nayebi, "Separating convolutive post non-linear mixtures", *Proceeding of ICA 2001*, San Diego (California, USA), December 2001, pp. 138-143.
This paper addresses blind source separation in convolutive post non-linear mixtures. In these mixtures, the sources are first mixed convolutively, and then distorted by nonlinear transforms due to the sensors. We show that the mutual information can be used as an independence criterion. The algorithm is based on the minimization of the mutual information, and uses multivariate score function estimation.

Publication 11 [**13**] M. Babaie-Zadeh, C. Jutten, K. Nayebi, "A geometric approach for separating post nonlinear mixtures", *Proceeding of EUSIPCO 2002*, Toulouse (France), September 2002, Vol. 2, pp. 11-14.
A geometric method for separating 2 sources from 2 PNL mixtures is presented. The main idea is to find two compensating nonlinearities able to transform the observations so that the joint plot of observed data is transformed to a plot contained in a parallelogram. It then results to a linear mixture which can be separated by any linear source separation algorithm. An indirect result of the paper is another separability proof of PNL mixtures for bounded sources for 2 sources and 2 sensors.

Publication 12 [**2**] S. Achard, D.T. Pham, C. Jutten, "Quadratic Dependence Measure for Nonlinear Blind source Separation", *Proc. of ICA 2003 Conference*, Nara (Japan), pages 263–268, April 2003.
This work focuses on a quadratic measure of dependence used to solve the problem of blind source separation. After defining it, we show some links with other quadratic measures used by Feuerverger and Rosenblatt. We develop a practical way for computing the criterion, which leads to a new method for blind source separation in nonlinear mixtures. It consists in first estimating the theoretical quadratic measure, then computing its relative gradient, finally minimizing it through a gradient descent procedure. Some examples illustrate the method in the most nonlinear case.

Publication 13 [**14**] M. Babaie-Zadeh, C. Jutten, K. Nayebi, "Minimization-Projection (MP) Approach for Blind Source Separation in Different Mixing Models", *Proceeding of ICA 2003 Conference, Nara (Japan)*, April 2003, pp. 1083-1088.
In this paper, a new approach for blind source separation is presented. This approach is based on minimization of the mutual information of the outputs using a non-parametric "gradient" of mutual information, followed by a projection on the parametric model of the separation structure. It is applicable to different mixing system, linear as well as nonlinear, and the algorithms derived from this approach are very fast and efficient.

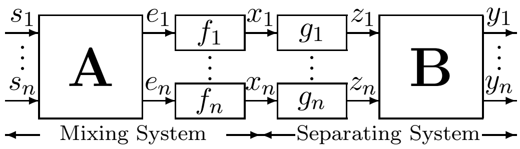

Figure 1: The mixing-separating system for PNL mixtures.

Publication 14 [51] A. Larue, C. Jutten, "Separation of Markovian sources in PNL mixtures", *Internal BLISS report*, September 2003.
Markovian source separation in post-nonlinear mixtures is the extension to PNL mixtures to Markovian approaches for linear mixtures, developed in the second year of the project. Using this approach [37], we minimize the conditional mutual information of the estimated sources y_i. We derive the estimation equation in the general case. However, for sake of simplicity, experimentally, we design the method only for nonlinear functions g_i's which are parameterized by one parameter θ_i or by a polynomial. Based on these equations, we design an algorithm which is much more efficient than algorithms considering iid signals, when sources are colored signals: one obtains an average (on 30 experiments) SNR equal to 60dB.

3 Separation methods for post-nonlinear mixtures

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

$$x_i(t) = f_i(\sum_{j=1}^{n} a_{ij} s_j(t)),\ i = 1, \dots, n \tag{3}$$

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

Besides its theoretical interest, this model belonging to the L-ZMNL[1] family suits perfectly for a lot of real-world applications. For instance, such models appear in sensors array processing [53], satellite and microwave communications [54], and in many biological systems [47].

3.1 Minimization of mutual information

Consider now BSS methods proposed for the simpler case of post-nonlinear mixtures (3). Taleb and Jutten have studied this case in several papers [65, 67, 66], and we start with a brief discussion of their results. A short overview of their studies can be found in [45], and the main results have been represented in [67].

The separation algorithm for the post-nonlinear mixtures (3) generally consists of two subsequent parts or stages:

[1] L stands for Linear and ZMNL stands for Zero-Memory Nonlinearity.

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 $\mathbf{y} = \mathbf{B}\mathbf{z}$ of the separating system are statistically independent (or as independent as possible).

Taleb and Jutten [67] 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 [38, 21]

$$\frac{\partial I(\mathbf{y})}{\partial \mathbf{B}} = -\mathrm{E}\{\boldsymbol{\psi}\mathbf{x}^T\} - (\mathbf{B}^T)^{-1} \tag{4}$$

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

$$\psi_i(u) = \frac{d}{du} \log p_i(u) = \frac{p_i'(u)}{p_i(u)} \tag{5}$$

Here $p_i(u)$ is the pdf of y_i and $p_i'(u)$ its derivative. In practice, the natural gradient algorithm [22, 20, 9] 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 [67]

$$\begin{aligned} \frac{\partial I(\mathbf{y})}{\partial \boldsymbol{\theta}_k} &= -\mathrm{E}\left\{ \frac{\partial \log \mid g_k'(\boldsymbol{\theta}_k, x_k) \mid}{\partial \boldsymbol{\theta}_k} \right\} \\ &\quad - \mathrm{E}\left\{ \sum_{i=1}^{n} \psi_i(y_i) b_{ik} \frac{\partial g_k(\boldsymbol{\theta}_k, x_k)}{\partial \boldsymbol{\theta}_k} \right\} \end{aligned} \tag{6}$$

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 [67], 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 (5) [67]. The score functions (5) must be estimated adaptively from the output vector $\mathbf{y}$. Several alternative ways to do this are considered in [67]. The first approach is to estimate the pdf, and then compute using differentiation the score function. Pdf estimation based on the Gram-Charlier expansion [24, 38] 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 [66].

Quadratic Dependence Measure for Nonlinear Blind Sources Separation: At INPG, we have considered and extended a dependence measure introduced by Eriksson *et al.*[27], which we call quadratic dependence measure. The advantage of this measure is its flexibility since one can choose a kernel with few constraints with a large bandwidth. Further, it can be easily estimated and we also obtain explicit formula for calculating the gradient of the estimated criterion. We have applied our method to the post nonlinear mixture model and have obtained satisfactory results [2].

Minimization-Projection algorithms During the second year, INPG derived the differential of the mutual information (MI), and showed that the stochastic gradient of the mutual information is nothing but the difference of marginal and joint score functions. This theoretical result leads to a new class of very efficient algorithms. Usually, given observations $\mathbf{x}$, source separation consists in estimating a mapping $\mathcal{G}$ so that $\mathbf{y} = \mathcal{G}(\mathbf{x})$ are independent; then, the mapping $\mathcal{G}$ is estimated so that the gradient of the mutual information $I(\mathbf{y})$ with respect to the parameters of $\mathcal{G}$ is equal to zero. The convergence of the algorithm strongly depends on the model complexity, which can for instance generate bad local minima. For avoiding this drawback, we share the algorithm in 2 stages:

- a minimization stage, in which one computes the data $\mathbf{y}$ so that they are as independent as possible, by using the differential of the MI (without assuming any model)
- a projection stage, in which one estimates the model parameters θ which fit the best to the data, according to a least square error criterion $J(\theta) = E[(\mathbf{y} - \mathcal{G}_\theta(\mathbf{x}))^2]$.

The method has been implemented for various models, especially post non-linear mixtures [15, 14], and points out faster convergence with respect to one-step algorithms.

Separability of bounded sources in PNL mixtures. A new separability theorem for PNL mixtures has been proposed for bounded sources, based on geometrical properties. For such sources, the joint distribution, after linear mixture, is a parallelogram. Then, restricting the mapping $\mathcal{H}$ to component-wise nonlinearity (PNL case: $\mathcal{H}_i(\mathbf{x}) = \mathbf{h_i}(\mathbf{x_i})$), we can prove that such a mapping transforms a parallelogram into another parallelogram only in the affine case: $h_i(x_i) = ax_i + b$. This result also suggests a new class of algorithm since the nonlinearity can be estimated without using an independence criterion, but exploiting the above geometrical property. Details on the separability theorem and on algorithm can be found in [11, 13].

Separation of Markovian sources in NL mixtures. Source separation in nonlinear case is, in general, impossible, since there exist many mappings with non-diagonal Jacobian matrices preserving the independence. At INPG, we wonder if the time structure (based on a Markovian model) of the sources can reduce this indeterminacy. In particular, we show [37] that the classical example used in literature for demonstrating the nontrivial non-separability of the non-linear mixtures can be rejected by taking into account the temporal correlation of the sources [37]. This result, although it does not insure that time dependent sources are separable in any nonlinear mixtures, points out on interest in taking into account prior informations for restricting the solution space.

Markovian source separation in post-nonlinear mixtures is the extension to PNL mixtures to Markovian approaches for linear mixtures, developed in the second year of the project. Using this approach [37], we minimize the conditional mutual information of the estimated sources y_i. We derive the estimation equation in the general case. However, for sake of simplicity, experimentally, we design the method only for nonlinear functions g_i's which are parameterized by one parameter θ_i. So the problem of the estimation of g_i, which canceled the nonlinear distortion f_i, is equivalent to the estimation of one parameter θ_i. Then, the separation problem consists in estimating the parameter vector $\Theta = [\theta_1, \ldots, \theta_N]$ and the separation matrix $\mathcal{B}$. The gradient of the criterion with respect to $\mathcal{B}$ is similar to the linear case replacing the observations with the nonlinear blocks outputs (after compensation by g_i). In [51], we derive the criterion with respect to the parameter θ_i of the nonlinear blocks:

$$\begin{aligned} \frac{\partial I(\mathcal{B}, \Theta)}{\partial \theta_i} &= -E\left[\frac{\partial^2 g_i(\theta_i, x_i(t))}{\partial x_i(t) \partial \theta_i} \left(\frac{\partial g_i(\theta_i, x_i(t))}{\partial x_i(t)}\right)^{-1}\right] + \\ & E\left[\sum_{j=1}^{N} b_{ji} \sum_{l=0}^{q} \psi_{y_j}^{(l)}(y_j(t)|y_j(t-1), \ldots, y_j(t-q)) \frac{\partial g_i(\theta_i, x_i(t-l))}{\partial \theta_i}\right] \end{aligned} \tag{7}$$

This expression also used the estimation of conditional score function $\psi_{y_j}^{(l)}(.)$, and the others terms are calculated with the explicit function $g_i(\theta_i, x_i(t))$. Initialization of the parameter θ_i is computed with a gaussianity criterion [64]. At each iteration of the algorithm, we compute the new separation matrix $\mathcal{B}$ by a relative gradient and the new nonlinear parameters θ_i by a natural gradient. With this algorithm, the separation of an instantaneous post-nonlinear mixtures of two colored Gaussian sources is very efficient: the average (on 30 experiments) SNR is 60dB, with a standard deviation of 20dB.

3.2 Linearization methods: ACE and Gaussianization

Another approach by Ziehe et al. [81] uses powerful linearization techniques in order to enable the subsequent application of standard linear ICA/BSS techniques for source separation in the PNL case.

First, the alternating conditional expectation (ACE) method of non-parametric statistics has been proposed for approximate inversion of the post-nonlinear functions f_i in (3). In a second step, a linear BSS method, e.g. based on temporal decorrelation (see [17, 84, 81]), is used for recovering the source signals. Independently, Solé et al. [64] and Ziehe et al. [82] 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

$$\hat{g}_i = \Phi^{-1} \circ F_{X_i} \tag{8}$$

Here F_{X_i} is the cumulative distribution function of the random variable X_i, and Φ is the cumulative Gaussian distribution. For a more detailed description and a comparison of those methods see the attached paper [83].

3.3 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 [68] as well as parametric [63] algorithms based on minimization of mutual information rate [25]. A similar problem appearing in satellite communications has been solved using Monte Carlo Markov Chain (MCMC) simulation methods [62].

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

$$x_i(k) = f_i([A(z)]s(k)),\ i = 1, \ldots, n \tag{9}$$

Some separation algorithms based on the generalization of mutual information minimization for random processes have been proposed in [12, 14].

PNL methods for designing smart chemical sensor array. This work, which was not intended in the project, has been done in cooperation with UPC (Barcelona), during a 4-month visit in INPG of a PhD student (Guillermo Bedoya) working in the framework of another European project (SEWING, IST-2000-28084). It is however reported here since it is an interesting application of PNL.

Ion-selective field effect transistor (ISFET) are used as chemical sensors for biomedical and environmental applications, since ion activities at the interface between the electrolyte solution and the ISFET membrane vary the gate voltage (and consequently the drain current), according to:

$$V_G = V_{Ref} + \frac{RT}{n_i F} T \ln(s_i + \sum_j K_{ij} s_j^{\frac{z_i}{z_j}}) \tag{10}$$

where R is the gas constant, T is the temperature, n_i is the charge of the measured ion, F is the Faraday constant, s_i is the activity of ion i, K_{ij} is the selectivity coefficient, z_i is the charge of the ion i, and s_i is the ion activity of ion i. For ions with the same charge, Eq. (10) is a PNL mixture, with known nonlinearity and unknown parameters. Then, for recovering the ions activity, one can estimate a PNL separation model, with exp as nonlinearities followed by a mixing matrix. Blind separation can be achieved by minimizing mutual information. Preliminary results [16] are promising and suggest another way for designing accurate ISFET sensor exploiting sensor variability.

4 Separation methods for general nonlinear mixtures

4.1 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 [50], 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) [38, 21]. 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. We shall not proceed deeper into the theory of variational Bayesian learning here, because this has been done already in the deliverable D17 [70] of this project as well as in the publication [76] included into the deliverable D21 [46]. Connections of variational Bayesian ensemble learning to methods and theory developed for efficient coding of information have been explored in [35].

Variational Bayesian methods were first applied to standard linear ICA and BSS in [10, 48], and several research groups have since then used Bayesian approaches to handle various blind problems for linear models; see [28, 38, 56, 76] and the references therein. H. Valpola (earlier Lappalainen) and his co-authors have introduced at HUT several methods based on Bayesian ensemble learning for blind estimation and separation in nonlinear mixture (data) models.

In these methods, the nonlinear model in (1) is modified somewhat so that it contains also an additive noise term:

$$\mathbf{x}(t) = \mathbf{f}(\mathbf{s}(t)) + \mathbf{n}(t) \tag{11}$$

The nonlinear mapping $\mathbf{f}$ in (11) is modeled using a multilayer perceptron (MLP) network [32] with one nonlinear hidden layer:

$$\mathbf{f}(\mathbf{s}) = \mathbf{B}\tanh(\mathbf{A}\mathbf{s} + \mathbf{a}) + \mathbf{b} \tag{12}$$

The mapping $\mathbf{f}$ is thus parametrized using weight matrices $\mathbf{A}$ and $\mathbf{B}$ as well as the bias vectors $\mathbf{a}$ and $\mathbf{b}$ of the hidden layer and output layer of the MLP network, respectively. The regularization needed in nonlinear BSS is achieved by choosing the model $\mathbf{f}(\mathbf{s})$ and sources $\mathbf{s}(t)$ that have most probably generated the observed data $\mathbf{x}(t)$ [70, 76].

Assuming that the source signals $\mathbf{s}$ at the input layer of the MLP network (12) have simple Gaussian distributions, one obtains a nonlinear principal component analysis (PCA) solution called nonlinear factor analysis (abbreviated NFA, or NLFA in some of our papers) [49, 69, 71, 38]. The NFA solution can usually model quite well the nonlinear mixtures (observed data) $\mathbf{x}(t)$, but it does not yet provide estimates of the independent source signals, because the sources $\mathbf{s}(t)$ 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 variational Bayesian ensemble learning, but using now a more sophisticated mixture-of-Gaussians model for the sources [49, 69, 71, 38]. It is well known that suitable mixtures of Gaussian distributions are able to model with sufficient accuracy any source distributions [18]. This method is called Nonlinear Independent Factor Analysis (NIFA).

The NFA and NIFA methods were first introduced in [49], and a more principled theoretical derivation has been presented in [69]. Experimental results with artificially generated data have been presented in [38, 49, 71], showing that the NFA method followed by linear ICA as well as the NIFA method are able to approximate pretty well the true sources. These methods have been applied also to real-world data sets, including 30-dimensional pulp data [38, 49, 71] and speech data [77], 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 [75]. The developed NDFA (Nonlinear Dynamic Factor Analysis) method is presented thoroughly in [76], and the results obtained thus far have been summarized in [77]. This paper has been included into this deliverable. The MATLAB codes for the NFA and NDFA methods are available at the www site [74].

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

$$\mathbf{x}(t) = \mathbf{f}(\mathbf{s}(t)) + \mathbf{n}(t) \tag{13}$$

$$\mathbf{s}(t) = \mathbf{g}(\mathbf{s}(t-1)) + \mathbf{m}(t) \tag{14}$$

In the latter equation (14), $\mathbf{g}(t)$ 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 (13). 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 (13)–(14) is discussed in detail in [76].

Many real-world data sets can be appropriately described as nonlinear dynamic systems such as (13)–(14), 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 [23], where the authors have considered state-space models and a hyper radial-basis function (RBF) network [32] for nonlinear mixtures. However, the method introduced in [23] is not completely blind, because it partly resorts to supervised learning.

In experiments with difficult chaotic data [75, 76], the NDFA method performed excellently, outperforming for example the prediction results given by nonlinear autoregressive modeling learned by standard back-propagation (see [32], Chapter 15) by an order of magnitude. The NDFA method has been applied also to BSS of biomedical MEG data in [57], and it provided clearly better results than standard linear ICA. In this application, the static data model (13) was a standard linear ICA model, but the dynamic model (14) 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 [42, 77]. The method performed again much better than the compared state-of-the-art techniques of change detection. The results of this study will be presented in detail in the forthcoming journal paper [43].

A problem in particular with the NDFA method but also with the NFA and NIFA methods is that their computational load remains 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 was developed in [79].

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 [73, 72]. In the publication [78] included into this deliverable we have tested it in nonlinear BSS by constructing a hierarchical nonlinear data model from the blocks. The resulting method is called hierarchical nonlinear factor analysis (HNFA). The HNFA method gives for artificial data slightly worse results [78] than the NFA method followed by linear ICA or the NIFA method, but preliminary experiments with real-world speech data are quite encouraging. Further properties of the HNFA method in the task of reconstructing missing values have been studied in the conference paper [55] attached to this deliverable.

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 [41]. Bayesian ensemble learning can be accelerated also by applying an improved updating scheme for the parameters to be estimated; this has been studied in the papers [33, 36].

4.2 Kernel-based methods

Kernel-based learning has become a popular technique recently (e.g. [80, 26, 60, 19, 61, 52]). The basic idea of "kernelizing" (see [61]) allows to construct very powerful nonlinear variants of existing linear scalar product based algorithms by mapping the data $\mathbf{x}[t]$ ($t = 1, \ldots, T$) implicitly into some kernel feature space $\mathcal{F}$ through some mapping $\Phi : \Re^n \rightarrow \mathcal{F}$. Performing a simple linear algorithm in $\mathcal{F}$, then corresponds to a nonlinear algorithm in input space: in other words a linear blind source separation in $\mathcal{F}$ would give rise to a nonlinear BSS algorithm in input space. All can be done efficiently and never directly but implicitly in $\mathcal{F}$ by using the kernel trick $\mathbf{k}(\mathbf{a}, \mathbf{b}) = \Phi(\mathbf{a}) \cdot \Phi(\mathbf{b})$. However, a straight forward application of the kernel trick to BSS fails for two reasons: applying a linear BSS algorithm in feature space will not necessarily identify the sought-after signals, since there are very likely directions that are also independent but higher-order versions of the original signals, and secondly, in principle, the BSS algorithm has to be applied, after kernelizing, to a T dimensional problem which is numerically neither stable nor tractable.

In our contribution (see [30, 31] we added a new aspect which enabled us to first employ successfully kernel-based methods for nonlinear BSS: since typically the data forms a lower dimensional subspace in $\mathcal{F}$, even much lower than T dimensional, we apply first a dimension reduction step before applying the linear BSS algorithm. We therefore propose a mathematical construction—very much inspired by reduced set methods (e.g. [58])—that allows us to adapt to the intrinsic data dimension. In the next step an orthonormal basis of this low dimensional sub-manifold is constructed which eventually makes the computations of a subsequent BSS algorithm tractable. The subtle difference to reduced set techniques is that we do not aim to construct a low dimensional basis for a good classification, rather we aim for an efficient, i.e. low dimensional description of the data in $\mathcal{F}$. Note, that we use a BSS algorithm that is based on second order temporal decorrelation (see [84, 17]) which is an essential building block of our algorithm. Lastly, the sources of interest are automatically selected after the BSS step. These ingredients give rise to an algorithmic solution, called kTDSEP, that is mathematically elegant and efficient with high performance, as the experiments on nonlinear mixtures of artificially generated signals and various sound signals (for details see attached paper).

4.3 Combination of kernel PCA and NFA

In the nonlinear factor analysis (NFA) method [49, 69, 77] developed in HUT, an appropriate initialization is essential for getting good results. This results from the flexibility of the multilayer perceptron (MLP) network [32] used to model the nonlinear mixing process, and from the general ill-posed nature of the nonlinear blind source separation problem [39, 38, 44]. In the original implementation, initialization of the sources is carried out by computing a desired number of first linear principal components of the data and then fixing the sources to those values for some time while the MLP network is adapted.

In co-operation with HUT and FhG, it was studied whether the kernel PCA method [32, 59] developed earlier at FhG could be used to improve the initialization of the NFA method. Kernel PCA (KPCA) [59] is a nonlinear generalization of linear principal component analysis (PCA). It works by mapping the original data space nonlinearly to a high dimensional feature space and performing PCA in that space. With the kernel approach this can be done in a computationally efficient manner.

The results of this joint study are reported in detail in [34]. They show that with a suitable kernel PCA initialization, both the signal-to-noise ratios of the separated sources and the convergence speed of the NFA indeed improved clearly. This effect was more pronounced for the artificially generated data consisting of nonlinear mixtures of four sub-Gaussian and four super-Gaussian sources, while for real-world speech data the improvement was smaller. A problem with using kernel PCA for initialization and in general is that the choice of the nonlinear kernel requires care. If the kernel is chosen poorly, the results can be even poorer than when using linear PCA initialization [34].

4.4 Comparison of PNL

In a joint study of Helsinki University of Technology (HUT) and Institut National Polytechnique de Grenoble (INPG), two different approaches for solving the nonlinear BSS problem were compared for post-nonlinear mixtures. The Bayesian nonlinear factor analysis method (abbreviated NFA or NLFA) developed at HUT [49, 69, 77] can be applied to general nonlinear mixtures, while the post-nonlinear (PNL) method developed at INPG [67, 66] requires that the observed mixtures are post-nonlinear (PNL) in nature. That is, the mixtures are modeled as linear mixtures which have undergone some unknown nonlinear distortion. The results of this comparison have been presented in more detail in the attached publication [40].

The following conclusions were drawn from results of the comparison:

- The INPG methods are clearly superior for post-nonlinear mixtures having the same number of sources and observed mixtures, provided that all the post-nonlinearities are invertible.
- The performance of the INPG methods degrades when the number of mixtures exceeds the number of sources, but they still outperform the more general Bayesian methods.
- In general, the Bayesian methods developed at HUT may be more preferable in high-dimensional post-nonlinear BSS problems where the number of sources is small.
- The Bayesian methods are able to separate post-nonlinear mixtures with non-invertible post-nonlinearities while the existing INPG methods cannot do this.

This preliminary study showed the benefit of exploiting the additional information of more observations than sources, especially in the nonlinear mixtures case. In this case, globally invertible PNL mixtures, but with non-invertible component-wise nonlinearities, can be identified and sources can be separated, which is a new and interesting result [40]. It also emphasizes on the relevance (now well known) of a pertinent choice of the separation structure for improving the performance. Especially, PNLFA (i.e. NLFA with structural constraints suited to PNL) provides interesting results, much better than NLFA without constraint. Moreover, non Bayesian PNL approaches, whose separation structure is not suited to non-invertible PNL mixtures, could be extended based on PNLFA separation structure.

4.5 Mutual information based method

Mutual information is a measure of statistical dependence with a well understood meaning and with several desirable properties. It is therefore a natural choice as an objective function for ICA. Estimating the mutual information in the ICA context is not straightforward, however.

It is known that the popular INFOMAX method of linear ICA/BSS can be viewed as minimizing the mutual information of the extracted components, if its output nonlinearities are chosen as the cumulative probability functions (CPFs) of the sources. The MISEP method – which was developed mostly within the BLISS project – extends INFOMAX in two directions:

- By replacing the linear unmixing block with a generic nonlinear one (e.g. an MLP), thus allowing the processing of nonlinear mixtures.
- By replacing the fixed nonlinearities of INFOMAX with adaptive ones, that maintain online estimates of the CPFs of the components that are being estimated.

The resulting method performs nonlinear ICA using a minimum mutual information criterion. In practice it optimizes a nonlinear network with a specialized architecture, by maximizing its output entropy. MISEP has been shown, within the BLISS project, to be able to perform ICA and BSS in moderately nonlinear mixtures of random noise with various statistical distributions, as well as in nonlinear mixtures of speech signals. Nonlinear BSS was achieved due to the smoothness-related regularization that was performed (implicitly in most cases, but also explicitly in some ones).

INESC-ID had to withdraw from the BLISS project due to reasons internal to INESC-ID itself, but the development of the MISEP method proceeded. We summarize here two results obtained during this later period, because we think that they are relevant for an assessment of the usefulness of the method:

- The method was shown to be able to separate nonlinear mixtures of up to 10 sources, with a moderate increase in processing time.
- The method has been successfully used to separate nonlinear mixtures of images, occurring in a real-life application.

The publication [7] gives a full description of the method, as well an account of several experimental results (the results on the image separation application are subject to some confidentiality constraints, and are not published yet). Other publications on the MISEP method are [3, 4, 5, 6, 8].

5 Concluding remarks

In general, for nonlinear mixtures, a decomposition into independent components does not ensure recovery of the underlying sources. However, we showed that proper regularization strategies allow to solve the nonlinear BSS problem in many cases. We achieved the nonlinear source separation with a number of different approaches like post-nonlinear models, kernel-based methods and variational Bayesian methods. While those methods have been developed, we made much effort to compare the results and to pursue combined approaches. Examples are the joint work between HUT and INPG for extending the ensemble learning methods for post-nonlinear mixtures an the joint work between FhG and HUT for using kernel-PCA for initializing the Bayesian methods.

The different approaches related to the nonlinear ICA/BSS task have also been presented and discussed at major conferences: first, at the ICA 2003 workshop in the invited special session on nonlinear ICA and BSS organized by Christian Jutten and Juha Karhunen, and in their invited paper "Advances in Nonlinear Blind Source Separation", and second, at the NIPS 2002 workshop: "ICA and beyond" organized by Stefan Harmeling, Erkki Oja, Luis, Almeida and Dinh-Tuan Pham, Both events showed the leading role of the BLISS group in this field of research.

References

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

[2] S. Achard, D. T. Pham, and Ch. Jutten. Quadratic dependence measure for nonlinear blind source separation. In *Proceeding of ICA 2003 Conference*, pages 263–268, Nara, Japan, April 2003.

[3] 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)*, pages 117–122, Lake Louise, Canada, October 2000.

[4] L. Almeida. ICA of linear and nonlinear mixtures based on mutual information. In *Proc. 2001 Int. Joint Conf. on Neural Networks*, Washington, D.C., 2001.

[5] L. Almeida. MISEP – an ICA method for linear and nonlinear mixtures, based on mutual information. In *Proc. 2002 Int. Joint Conf. on Neural Networks*, Honolulu, Hawaii, 2002.

[6] L. Almeida. Faster training in nonlinear ICA using MISEP. In *Proc. Int. Worksh. Independent Component Analysis and Blind Signal Separation*, pages 113–118, Nara, Japan, 2003.

[7] L. Almeida. Linear and nonlinear ICA based on mutual information – the MISEP method. *Signal Processing, Special issue on ICA*, 2003. In press.

[8] L. Almeida. MISEP – linear and nonlinear ICA based on mutual information. *Journal of Machine Learning Research, Special issue on ICA*, 2003. In press.

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

[10] H. Attias. Independent factor analysis. *Neural Computation*, 11(4):803–851, 1999.

[11] M. Babaie-Zadeh. *On blind source separation in convolutive and nonlinear mixtures.* PhD thesis, INPG, Grenoble, France, September 2002.

[12] 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)*, pages 138–143, San Diego (California, USA), 2001.

[13] M. Babaie-Zadeh, C. Jutten, and K. Nayebi. A geometric approach for separating post nonlinear mixtures. In *Proceedings of EUSIPCO 2002*, volume 2, pages 11–14, Toulouse, France, September 2002.

[14] M. Babaie-Zadeh, C. Jutten, and K. Nayebi. A minimization-projection approach for separating post nonlinear mixtures. In *Proc. Int. Conf. on Independent Component Analysis and Signal Separation (ICA2003)*, pages 1083–1088, Nara, Japan, April 2003.

[15] M. Babaie-Zadeh, J. Solé-Casals, and C. Jutten. Blind inversion of wiener system using a minimization-projection (mp) approach. In *Proceeding of ICA 2003 Conference*, pages 681–686, Nara, Japan, April 2003.

[16] G. Bedoya, C. Jutten, S. Bermejo, and J. Cabestany. Improving semiconductor-based chemical sensor arrays using advanced algorithms for blind source separation. In *Proc. of SIcon/04*, New Orleans, USA, January 2004. Submitted to SIcon/04.

[17] A. Belouchrani, K. Abed Meraim, J.-F. Cardoso, and E. Moulines. A blind source separation technique based on second order statistics. *IEEE Trans. on Signal Processing*, 45(2):434–444, 1997.

[18] C. Bishop. *Neural Networks for Pattern Recognition.* Clarendon Press, 1995.

[19] C.J.C. Burges. A tutorial on support vector machines for pattern recognition. *Knowledge Discovery and Data Mining*, 2(2):121–167, 1998.

[20] J.-F. Cardoso and B. Laheld. Equivariant adaptive source separation. *IEEE Trans. on Signal Processing*, 44(12):3017–3030, 1996.

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

[22] A. Cichocki, R. Unbehauen, and E. Rummert. Robust learning algorithm for blind separation of signals. *Electronics Letters*, 30(17):1386–1387, 1994.

[23] 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)*, pages 99–104, Aussois, France, 1999.

[24] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, 1994.

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

[26] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines.* Cambridge University Press, Cambridge, UK, 2000.

[27] J. Eriksson, A. Kankainen, and V. Koivunen. Novel characteristic function based criteria for ICA. In *Proceeding of ICA 2001 Conference*, San Diego, Ca, USA, December 2003.

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

[29] S. Harmeling, A. Ziehe, M. Kawanabe, and K.-R. Müller. Kernel-based nonlinear blind source separation. Technical Report 1/2002, European Commission Research Project BLISS (Blind Source Separation and Applications), IST-1999-14190, 2002.

[30] S. Harmeling, A. Ziehe, M. Kawanabe, and K.-R. Müller. Kernel feature spaces and nonlinear blind source separation. In T.G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems*, volume 14, pages 761–768. MIT Press, 2002.

[31] S. Harmeling, A. Ziehe, M. Kawanabe, and K.-R. Müller. Kernel-based nonlinear blind source separation. *Neural Computation*, 15:1089–1124, 2003.

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

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

[34] A. Honkela, S. Harmeling, L. Lundqvist, and H. Valpola. Using kernel PCA for initialization of nonlinear factor analysis. Technical Report 1/2003, European Commission Research Project BLISS (Blind Source Separation and Applications), IST-1999-14190, 2003.

[35] A. Honkela and H. Valpola. Variational learning and bits-back coding: an information theoretic view to bayesian learning. 2003. Submitted to a journal.

[36] A. Honkela, H. Valpola, and J. Karhunen. Accelerating cyclic update algorithms for parameter estimation by pattern searches. *Neural Processing Letters*, 17(2):191–203, 2003.

[37] S. Hosseini and C. Jutten. On the separability of nonlinear mixtures of temporally correlated sources. *IEEE Signal Processing Letters*, 10(2):43–46, February 2003.

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

[39] A. Hyvärinen and P. Pajunen. Nonlinear independent component analysis: Existence and uniqueness results. *Neural Networks*, 12(3):429–439, 1999.

[40] A. Ilin, A. Honkela, S. Achard, and C. Jutten. The comparison of the blind source separation methods developed at HUT and INPG. Technical Report 2/2003, European Commission Research Project BLISS (Blind Source Separation and Applications), IST-1999-14190, 2003.

[41] 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)*, pages 915–920, Nara, Japan, April 2003.

[42] 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)*, pages 704–709, San Diego, USA, December 2001.

[43] A. Ilin, H. Valpola, and E. Oja. Nonlinear dynamical factor analysis for state change detection. *IEEE Trans. on Neural Networks*, 2003. Submitted (under revision).

[44] C. Jutten and J. Karhunen. Advances in nonlinear blind source separation. In *Proc. of the 4th Int. Symp. on Independent Component Analysis and Blind Signal Separation (ICA2003)*, pages 245–256, Nara, Japan, April 1-4 2003. Invited paper in the special session on nonlinear ICA and BSS.

[45] 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)*, pages 15–26, Helsinki, Finland, 2000.

[46] J. Karhunen, F. Meinecke, and H. Valpola. Final technical report on BSS models and methods for non-independent BSS. Technical Report Deliverable D21, European Commission Research Project BLISS (Blind Source Separation and Applications), IST-1999-14190, September 2003.

[47] M. Korenberg and I. Hunter. The identification of nonlinear biological systems: LNL cascade models. *Biol. Cybernetics*, 43(12):125–134, December 1995.

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

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

[50] H. Lappalainen and J. Miskin. Ensemble learning. In M. Girolami, editor, *Advances in Independent Component Analysis*, pages 75–92. Springer-Verlag, Berlin, 2000.

[51] A. Larue and C. Jutten. Separation of markovian sources in PNL mixtures, September 2003.

[52] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf. An introduction to kernel-based learning algorithms. *IEEE Transactions on Neural Networks*, 12(2):181–201, 2001.

[53] 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*, pages 666–671, Santa Clara, CA, USA, April 1998.

[54] S. Prakriya and D. Hatzinakos. Blind identification of LTI-ZMNL-LTI nonlinear channel models. *IEEE Trans. on Signal Processing*, 43(12):3007–3013, December 1995.

[55] T. Raiko, H. Valpola, T. Östman, and J. Karhunen. Missing values in hierarchical nonlinear factor analysis. In *Proc. of Int. Conf. on Artificial Neural Networks and Neural Information Processing (ICANN/ICONIP 2003)*, pages 185–188, Istanbul, Turkey, 2003.

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

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

[58] B. Schölkopf, S. Mika, C.J.C. Burges, P. Knirsch, K.-R. Müller, G. Rätsch, and A.J. Smola. Input space vs. feature space in kernel-based methods. *IEEE Transactions on Neural Networks*, 10(5):1000–1017, September 1999.

[59] B. Schölkopf, A. Smola, and K.-R. Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5):1299–1319, 1998.

[60] B. Schölkopf and A.J. Smola. *Learning with Kernels*. MIT Press, Cambridge, MA, 2002.

[61] B. Schölkopf, A.J. Smola, and K.-R. Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10:1299–1319, 1998.

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

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

[64] J. Solé-Casals, M. Babaie-Zadeh, Ch. Jutten, and D. T. Pham. Improving algorithm speed in PNL mixture separation and wiener system inversion. In *Proceeding of ICA 2003 Conference*, pages 639–644, Nara, Japan, April 2003.

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

[66] A. Taleb and C. Jutten. Batch algorithm for source separation in post-nonlinear mixtures. In *Proc. First Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, pages 155–160, Aussois, France, 1999.

[67] A. Taleb and C. Jutten. Source separation in post-nonlinear mixtures. *IEEE Trans. on Signal Processing*, 47(10):2807–2820, 1999.

[68] A. Taleb, J. Sole, and C. Jutten. Quasi-nonparametric blind inversion of Wiener systems. *IEEE Trans. on Signal Processing*, 49(5):917–924, 2001.

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

[70] H. Valpola. Technical report on Bayesian methods. Technical Report Deliverable D17, European Commission Research Project BLISS (Blind Source Separation and Applications), IST-1999-14190, June 2003.

[71] 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)*, pages 351–356, Helsinki, Finland, 2000.

[72] H. Valpola, M. Harva, and J. Karhunen. Hierarchical models of variance sources. *Signal Processing*, 2003. Tentatively accepted (under revision).

[73] H. Valpola, M. Harva, and J. Karhunen. Hierarchical models of variance sources. In A. Cichocki and N. Murata, editors, *Proc. of 4th Int. Symp. on Independent Component Analysis and Blind Source Separation (ICA2003)*, pages 83–88, Nara, Japan, April 2003.

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

[75] 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)*, pages 460–465, Honolulu, HI, USA, 2002.

[76] H. Valpola and J. Karhunen. An unsupervised ensemble learning method for non-linear dynamic state-space models. *Neural Computation*, 14(11):2647–2692, 2002.

[77] H. Valpola, E. Oja, A. Ilin, A. Honkela, and J. Karhunen. Nonlinear blind source separation by variational Bayesian learning. *IEICE Transactions (Japan)*, E86-A(3):532–541, March 2003.

[78] H. Valpola, T. Östman, and J. Karhunen. Nonlinear independent factor analysis by hierarchical models. In A. Cichocki and N. Murata, editors, *Proc. of 4th Int. Symp. on Independent Component Analysis and Blind Source Separation (ICA2003)*, pages 257–262, Nara, Japan, April 2003.

[79] 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)*, pages 710–715, San Diego, USA, 2001.

[80] V.N. Vapnik. *The nature of statistical learning theory.* Springer Verlag, New York, 1995.

[81] 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)*, pages 433–438, San Diego, USA, 2001.

[82] A. Ziehe, M. Kawanabe, S. Harmeling, and K.-R. Müller. Blind separation of post-nonlinear mixtures using gaussianizing transformations and temporal decorrelation. In *Proc. Int. Conf. on Independent Component Analysis and Signal Separation (ICA2003)*, pages 269–274, Nara, Japan, April 2003.

[83] A. Ziehe, M. Kawanabe, S. Harmeling, and K.-R. Müller. Blind separation of post-nonlinear mixtures using linearizing transformations and temporal decorrelation. *Journal of Machine Learning research, Special issue on ICA*, 2003. in press.

[84] A. Ziehe and K.-R. Müller. TDSEP - an efficient algorithm for blind separation using time structure. In L. Niklasson, M. Bodén, and T. Ziemke, editors, *Proceedings of the 8th International Conference on Artificial Neural Networks*, Perspectives in Neural Computing, pages 675–680, Berlin, 1998. Springer Verlag.