INPUT VARIABLE SELECTION USING INDEPENDENT COMPONENT ANALYSIS AND HIGHER ORDER STATISTICS

Andrew D. Back and Andrzej Cichocki

Brain Science Institute, RIKEN
The Institute of Physical and Chemical Research
2-1 Hirosawa, Wako–shi, Saitama 351-0198 Japan.

ABSTRACT

In real world problems of nonlinear model building there may be a number of inputs available for use. However, a common problem is that we do not know which inputs are necessary for the model. Previous methods have difficulties in coping with dependent inputs. In this paper, we propose a novel method of input variable selection based on independent component analysis and higher order cross statistics. Experimental results indicate that the method is capable of giving reliable performance with dependent inputs to nonlinear models.

1. INTRODUCTION

To construct nonlinear models of real world data, it is often necessary to distinguish between a number of possible inputs. We may have many inputs available for use yet a problem may exist in that we do not know how many or which inputs to the model are necessary. As the input dimensionality increases however, serious problems can arise. *Input variable selection* (IVS) is aimed at determining which input variables are required for a model.

We may differentiate between *model-free* and *model-based* methods [1]. In classical time series literature, a well known model-based test for IVS is the global F-test. Given a linear regression model

$$\hat{y} \quad = \quad \boldsymbol{\theta}^T \mathbf{x} \tag{1}$$

where $\boldsymbol{\theta}$ is an $m \times 1$ parameter vector and $\mathbf{x} = [x_1 \cdots x_m]^T$ is an $m \times 1$ input regression vector, the aim is to test the inputs $\{x_i\}$ to determine whether they improve prediction of the desired output y. In practice, the global F-test is implemented by means of analysis of variance (ANOVA).

More recently, model based approaches using nonlinear models such as neural networks have been proposed. The Automatic Relevance Detection (ARD) approach proposed by Neal [2] is based on Bayesian statistics. Intuitively, the aim of the ARD algorithm is to reduce all unnecessary weights in the network to zero. Another neural network approach using backward elimination was proposed by Moody and Utans [3]. In this case, input variables which have the smallest sensitivity with respect to the network output(s) are removed.

A model-free IVS method based on *mutual information* was proposed by Bonnlander [1]. The idea in this case is that the relevant inputs are found by estimating the mutual information between the inputs and the desired outputs. This approach performs significant computations in order to make a simple binary decision about the relevance of each input to the desired output. However, since some of the inputs may be dependent, for n input variables, it is necessary to test $2^n - 1$ subsets of inputs.

While many methods have been proposed, typically they all rely on some form of statistical test between the inputs and the desired output(s). Implicit in these methods is the assumption that the inputs observed are 'true' signals that can be used in the model. Assume we have a generating system given by

$$\mathbf{y}(t) \quad = \quad F_o\left(\mathbf{y}(t-1); \boldsymbol{\theta}\right) \tag{2}$$

where $\mathbf{y}(t) = [y_1(t) \cdots y_n(t)]^T$ and F_o is some static nonlinear function, possibly including a tapped delay line at the input. Such a model may represent a discretized set of coupled nonlinear differential equations for example. Suppose however, that we can write down this system in terms of the individual outputs as

$$y_i(t) \quad = \quad F_{oi}\left(\mathbf{y}_{ai}(t-1); \theta_i\right) \tag{3}$$

where $\mathbf{y}_{ai} \subseteq \mathbf{y}$ is a subset of the system outputs. One approach to determine $\mathbf{y}_{ai}$ is to test the independence between $y_i(t)$ and $y_j(t-1)$ $\forall i, j$.

In this paper, we consider a different problem, but one which may arise readily in practice. Here we consider the situation where we do not have access to the true outputs from a system at all, but only to some *mixture* of those signals. In such cases, we would like to model the observed signals using the available measured data, however conventional statistical tests will not necessarily reveal which inputs are required. Instead, the dependence between measured signals will simply indicate that all input variables may be required.

We propose that a method of *input variable transformation* (IVT) is required to ensure that the inputs applied to a model are made as independent as possible for the purpose of modelling multivariate time series. Such a method can be obtained using the recently developed procedure known as *independent component analysis* (ICA). While ICA has been applied to transforming data in the past [4], our application of ICA together with IVS appears to be new. In this paper, we aim to provide an understanding of the issues involved in selecting input variables and moreover, to indicate how ICA can be effectively applied to the problem.

The paper is organized as follows. A brief overview of ICA is given in Section 2. In Section 3 we describe the proposed methodology of input variable transformation using ICA. An example using the method is given in Section 4. Conclusions are given in Section 5.

2. INDEPENDENT COMPONENT ANALYSIS

Independent component analysis extends the method of principal component analysis (PCA) from decorrelating mixed signals (using second order statistics) to unmixing non-Gaussian signals such that they become statistically independent. Independent component analysis of instantaneously mixed signals is defined as follows. A multivariate time series $\mathbf{x}(t) = [x_1(t), x_2(t), ..., x_n(t)]'$ is obtained from a mixture of signals defined by

$$\mathbf{x}(t) \quad = \quad \mathbf{A}\mathbf{s}(t) \tag{4}$$

where $\mathbf{A}$ is the mixing matrix and $\mathbf{s}(t)$ is the vector of unobservable inputs. We seek to find a demixing matrix $\mathbf{W}$, where the output is an estimate of $\mathbf{s}(t)$. Hence we have

$$\begin{aligned} \mathbf{y}(t) &= \mathbf{W}\mathbf{x}(t) && (5) \\ &= \mathbf{W}\mathbf{A}\mathbf{s}(t) && (6) \end{aligned}$$

If $\mathbf{W} = \mathbf{A}^{-1}$, $\mathbf{y}(t) = \mathbf{s}(t)$ (here we ignore the problem of permutations and scaling factors) then the input signals will be obtained without error [5]. To find such a matrix $\mathbf{W}$, the assumptions made are that the sources $\{s_j(t)\}$ are statistically independent, at most one signal has a gaussian pdf, and that the signals are stationary.

Various algorithms for ICA have been proposed in the literature, see for example [6, 7, 8, 9, 10]. For the purposes of this paper, we assume the user selects an algorithm which is suited to the type of signals being separated.

3. ICA INPUT VARIABLE TRANSFORMATION

3.1. System Identification

The problem we consider in this case is as follows. Suppose there exists a system described by

$$y \quad = \quad F_o\left(\mathbf{s}_a; \boldsymbol{\theta}\right) \tag{7}$$

where F_o is some nonlinear mapping from a vector of inputs $\mathbf{s}_a = [s_1 \cdots s_p]'$ to an output y. Now, suppose that $\mathbf{s}_a$ is unobservable. Instead, we have available a vector of measurements $\mathbf{x} = [x_1 \cdots x_n]'$ which is a mixture of a superset of the unobservable measurements. Hence we have

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{8}$$

where $\mathbf{s} = [s_1 \cdots s_n]'$ $n > p$, $\mathbf{A}$ is a mixing matrix which may be rectangular, but for the present we will use a square $n \times n$ matrix. Therefore our model is:

$$\hat{y} = F\left(\mathbf{x}; \hat{\boldsymbol{\theta}}\right) \tag{9}$$

For input variable selection, it is normally assumed that there exists some subset of inputs $\mathbf{x}_a$ required by the model which can be used to accurately approximate the unknown system. In the past, there have been many input variable selection methods proposed to determine which inputs $\{x_i\}$to a model are required. However for equivalent model and system structures, since the true inputs are $\mathbf{s}_a$ not $\mathbf{x}_a$, there may not exist such a subset at all. If the true inputs $\mathbf{s}_a$ are not observable, then regardless of the variable selection tests applied, there does not necessarily exist any set of inputs $\mathbf{x}_a$ which is the correct set, apart from the whole set of inputs.

Hence we require a method of transforming $\mathbf{x}$ to an estimate of $\mathbf{s}_a$ and hence a form in which we can apply an input variable selection test. Whereas in the past only input variable selection methods have been considered, in this paper, we propose that a method of *input variable transformation* is also required. This ensures that the inputs applied to a model are independent (or as close as possible) and hence the input variable tests can proceed on each input independently. We propose that a method of finding $\hat{\mathbf{s}} = \mathbf{W}\mathbf{x}$ as required can be achieved using independent component analysis. This may seem to be a small advance in the use of ICA, however the issue of input variable selection is by no means trivial. Moreover, this is the first time that we are aware of, that ICA has been applied to IVS.

Consider a system as described in (7). The true inputs $\mathbf{s}_a$ are unobservable but instead we have available a vector of measurements $\mathbf{x}$. Since $\mathbf{s}_a \subseteq \mathbf{s}$, instead of the model given in (9), we seek a model

$$\begin{aligned} \hat{y} &= F\left(\hat{\mathbf{s}}_a; \hat{\boldsymbol{\theta}}\right) && (10) \\ \hat{\mathbf{s}}_a &= \mathbf{G}\hat{\mathbf{s}} && (11) \\ \hat{\mathbf{s}} &= \mathbf{W}\mathbf{x} && (12) \end{aligned}$$

where $\mathbf{W}$ is a demixing matrix to obtain an estimate of the original inputs $\hat{\mathbf{s}}$, and $\mathbf{G}$ is a sparse identity matrix which extracts only those signals required. Determining $\mathbf{G}$ and hence which elements of $\hat{\mathbf{s}}$ are required in the model can be obtained using any of the well known input variable selection methods. Determining $\mathbf{W}$ and hence the superset of unmixed inputs to the model, can achieved using any appropriate ICA algorithm [11].

3.2. Prediction

In contrast to the system identification case above, where there is a desired output available as in (7), the prediction case can be considered. Here the desired output signal can be generated using the one step ahead prediction of an input. In this manner we may consider prediction models.

3.2.1. Generating Models

We assume that there exists a general nonlinear dynamic system which generates a time series $\{y(t)\}$ given by

$$\mathbf{y}(t) = F_o\left(\mathbf{u}(t), \mathbf{y}_a(t-1)\right) \tag{13}$$

where $\mathbf{u}(t)$ is the driving input to the *system* which we are trying to predict, $\mathbf{y}(t) = [y_1(t) \cdots y_n(t)]^T$ are the multivariate outputs from the system (it is possible that $n = 1$) and $\mathbf{y}_a \subseteq \mathbf{y}$ is a subset of observed signals. In other words, we allow for the possibility that some of the outputs are fed back to the input of the model. We assume that the outputs $\mathbf{y}(t)$ are observable directly and are not subject to any mixing processes.

Note that implicit in the model F_o, we assume there exists the possibility of tapped delay lines occurring in exactly the same manner as a linear system. That is, if the input is $y(t)$, then the model may obtain as inputs, $y(t-1), y(t-2)...y(t-m)$.

When F_o is some highly nonlinear dynamic system and we cannot observe $\mathbf{u}(t)$, the problem of estimating or predicting $y(t)$ is difficult. There are a number of common simplifications and assumptions that are typically and often implicitly, applied to prediction models. A common assumption is that $\mathbf{u}(t)$ either does not exist, or that it plays a minor part in determining $y(t)$. Hence we may have a prediction model given by

$$\mathbf{y}(t) = F_o\left(\mathbf{y}(t-1)\right) \tag{14}$$

Hence the modelling task is to take the observed outputs and use them in a prediction model[1] given by

$$\hat{\mathbf{y}}(t) = F\left(\mathbf{y}(t-1)\right) \tag{15}$$

[1] It may appear that at that point all we achieve is a model of the input, however there are at least two major approaches that we may use for k-step prediction: (i) (classical) time delayed input training [12] and (ii) iterative prediction [13].

This form of model includes the usual linear prediction model based on an autoregressive (AR) structure, defined as

$$\hat{y}(t) = A(z)y(t) \tag{16}$$

$$A(z) = \sum_{i=1}^{n_a} a_i z^{-i} \tag{17}$$

Consideration of the generating models in this framework allows us to intelligently prescribe models for the prediction task. The next stage is to apply a method of input variable selection to determine the required inputs. We consider two major pertaining to this issue below.

3.2.2. Feedback Systems

Consider a feedback system of the form

$$y_i(t) = F_{oi}\left(\mathbf{y}(t-1)\right) \quad i = 1...n \tag{18}$$

where F_{oi} is a submodel contained in F_o. Such a model allows for the coupling of feedback systems. If such a model generates the observed data, what can be done to form appropriate prediction models?

We may begin by asking how strongly these systems are coupled. If the submodels are fully interconnected, then all outputs are fedback as inputs to all other submodels. If the submodels are not fully interconnected, then the system can be written as

$$y_i(t) = F_{oi}\left(\mathbf{y}_{ai}(t-1)\right) \quad i = 1...n \tag{19}$$

where $\mathbf{y}_{ai} \subseteq \mathbf{y}$ $i = 1...n$. This means we may use submodels of the form

$$\hat{y}_i(t) = F_i\left(\hat{\mathbf{y}}_{ai}(t-1)\right) \quad i = 1...n \tag{20}$$

In this situation, we need to apply a statistical test to each $y_i(t-1)$ to estimate the dependencies and hence the coupling between modules in the generating system. For nongaussian signals $\mathbf{y}$, an obvious approach is to use a test based on higher order statistics.

3.2.3. Dependent Observations

The problem of input variable selection can be clearly compounded by dependencies between the input variables. This may arise due to coupling between inputs, for example, due to coupled feedback systems or it may arise due to the observed data being mixed. In the past, this problem has been recognized (see for example [1]). The general approach has evidently been to use an exhaustive subset selection method to choose the best subset of input variables.

In the case where the observed data is simply a mixed version of the 'true' signals, we may observe $\mathbf{z}$, where

$$\mathbf{z} = \mathbf{B}\mathbf{y} \tag{21}$$

where $\mathbf{B}$ is a mixing matrix. In this case, there is clearly a dependence between each output y_i and so it is difficult to accurately estimate the sets $\mathbf{y}_{ai}$ $i = 1...n$. Therefore, at this stage, we may again apply ICA demixing to obtain

$$\tilde{\mathbf{y}} = \mathbf{Q}\mathbf{z} \tag{22}$$

where $\mathbf{Q}$ is a demixing matrix. Then the desired statistical IVS tests may be performed on $\tilde{y}_i$ to determine the true relationship between the modules. Since the demixing may not be perfect, it is necessary to ensure that the statistical test can determine those signals which are clearly independent.

Hence, our model is

$$\begin{aligned} \hat{\mathbf{z}} &= \hat{\mathbf{B}}\hat{\mathbf{y}} && (23) \\ \hat{y}_i(t) &= F_i\left(\tilde{\mathbf{y}}_{ai}(t-1)\right) \quad i = 1...n && (24) \\ \tilde{\mathbf{y}}_a &= \mathbf{G}\tilde{\mathbf{y}} && (25) \\ \tilde{\mathbf{y}} &= \mathbf{Q}\mathbf{z} && (26) \end{aligned}$$

3.3. An ICA-HOS IVT Algorithm

In this section we propose a method of overcoming the input variable selection problems identified previously. As shown above, input variable selection may suffer from the problem of dependent inputs which could be due to various reasons. In the past, it has been assumed that because of the dependencies between the inputs, the only feasible solution to determine the best input subset is to use exhaustive tests. In addition, there is a strong school of thought which dictates that model-based testing methods give better indications of the suitability of the input subset than model-free methods which further adds to the computational burden and difficulty of IVS.

Our approach is to apply ICA to preprocess the observed data. This in itself is not new. However, instead of simply filtering out the noisy data or removing artifacts on some arbitrary basis, it is suggested that ICA be used as a preprocessing method to obtain new input signals that are as independent as possible. If all the ICs are completely independent then we are satisifed also. However, even we if we can obtain groups of signals which are independent and some signals within the groups are dependent in some sense, then we can reduce the computational burden by only considering subsets within those groups. Thus, ICA is used as an integral part of IVS.

The particular IVS test we propose is based on the fourth order cross cumulants. This statistical measure can be used to establish the independence or otherwise of two non-Gaussian signals. The fourth order cross cumulant is defined as [14]

$$\begin{aligned} C_{abcd} &= E[abcd] - E[ab]\,E[cd] \\ &\quad -E[ac]\,E[bd] - E[ad]\,E[bc] \end{aligned} \tag{27}$$

where a, b, c, d are zero-mean input variables.

In practice, we find that taking a slice of the fourth order cumulants is sufficient to perform the test. Hence, for an input signal x, obtained as the output from the ICA algorithm, we compute the fourth order cross cumulant with the desired output y. For example, we have

$$\begin{aligned} C_{xyxx} &= E[xyxx] - E[xy]\,E[xx] \\ &\quad -E[xx]\,E[yx] - E[xx]\,E[yx] \\ &= E[xyxx] - 3E[xx]\,E[xy] \end{aligned} \tag{28}$$

The test is then

$$R_H = \begin{cases} 1 & C_{xyxx} > K \\ 0 & C_{xyxx} \leq K \end{cases} \tag{29}$$

where R_H is the test result and K is chosen to suit the level of precision required in determining the independence of the signals. The value of K is problem dependent and is chosen according to how the level of dependence that can be tolerated between the measured variables. In practice, we find that it is easy to select a suitable value.

To the best of our knowledge, this test has not been proposed in the literature previously. The reason for this is likely to be the fact that it depends on the inputs to be independent, hence without such a mechanism being available in the past to ensure this independence, such a test would have failed. The test provide a methods of avoiding the expensive model-based tests for IVS that have been very commonly used in the past, presumably as a means of overcoming the problem of dependent inputs. Thus, it can be observed that ICA provides a means of introducing a very simple, yet powerful IVS test which could not be used otherwise.

The proposed algorithm is summarized below.

ICA-HOS IVT Algorithm

1. Measure multivariate time series $\mathbf{z}$.
2. Perform ICA on $\mathbf{z}$ to obtain (26) and the estimated mixing matrix $\hat{\mathbf{B}} = \mathbf{Q}^{-1}$.
3. Apply the proposed HOS input variable selection test. Hence obtain $\mathbf{G}$ such that

$$\tilde{y}_{ai} = \begin{cases} \tilde{y}_i & R_H = 1 \\ 0 & R_H = 0 \end{cases} \tag{30}$$

where $R = 1$ indicates a positive outcome of the IVS test, that is, the variables are dependent, and $R = 0$ indicates a negative outcome of the IVS test, ie the variables are independent. This gives the required set of input variables.

Remarks.

1. ICA algorithms usually depend on the measured signals being non-Gaussian or at most one signal can be Gaussian. The reason for this is that the higher order statistics of Gaussian signals are zero and contain no additional information. On the other hand, non-Gaussian signals have non-zero higher order statistics and this can be exploited in ICA algorithms.

2. Some ICA algorithms require only second order statistics [15]. Since it was assumed in step 2 that the inputs are non-Gaussian, then we are able to use a HOS test on the data. However, provided the signals are separated into independent components in step 2, second order statistics could also be used.

3. If we assume that the ICs obtained are all fully independent, then it is no longer necessary to exhaustively consider all $2^n - 1$ subsets of the n input signals. In this case only n statistical tests per output signal are required which provide a substantially reduced computational burden. In addition, the robustness of the test results can also be expected to improve.

4. EXAMPLE

Here we give an example which demonstrates the ICA-HOS input variable selection method proposed in the paper. The example we choose is an artificial problem which nevertheless demonstrates the method.

Consider a set of four independent signals $\mathbf{y} = [y_1, y_2, y_3, y_4]'$ which are assumed to be generated by four independent systems, where

$$\begin{aligned} y_1 &= s_q(2\pi 29T) + 0.1N(0,1) \\ y_2 &= \sin(2\pi 300T + 6\cos(2\pi 60T)) \\ y_3 &= \cos(2\pi 153T + 3\cos(2\pi 37T)) \\ y_4 &= v(2\pi 93T) + w_e(1,2) \end{aligned} \tag{31}$$

s_q is a square-wave source, v is a sawtooth source and w_e is random noise with a Weibull distribution. Apart from some gaussian noise in one channel, the signals are nongaussian.

Now let us suppose that there is a nonlinear model F_o, in which

$$y_0(t) = F_0(\mathbf{y}_a) \tag{32}$$

$$\mathbf{y}_a = [y_{a1}, y_{a2}]' \tag{33}$$

$$F_{01} = \cos(y_{a1}) \tag{34}$$

$$F_{02} = y_{a2}^2 \tag{35}$$

Out of the four signals, only two are used to generate the output of interest. We observe $\mathbf{z}$ which is an $n \times 1$ vector of inputs to our model. The aim of the proposed method is to determine a transformation of $\mathbf{z}$ which will lead to a more accurate prediction.

Using the method proposed in this paper, we apply ICA to the measured signals $\mathbf{z}$ to obtain $\hat{\mathbf{y}}$, an estimate of $\mathbf{y}$. We use the fourth order cumulant test on input variables as described in Section 4. The cross-cumulants are taken between the input signals and desired output. High values imply dependence and that the signal is required. In this case, we select $K = 2$ as the threshold value.

Applying the IVS test to the raw inputs $\mathbf{z}$, we find that the fourth order cumulants values are:

$$C = [-15.7, 2.6, -5.8, -18.2] \tag{36}$$

It is readily apparent that the signals have varying degrees of independence with the desired output, but it is not clear whether the low values indicate whether those inputs are required or not. Based on the K threshold value chosen, it is indicated that four inputs are required.

The cross-cumulant values[2] obtained are

$$C = [-23.9, 12.2, 0.65, 1.06] \tag{37}$$

In this example, it is obvious that two signals have large values in comparison to the other two signals. The difference between values has widened, and the proposed ICA method correctly indicates the two signals required.

5. CONCLUSIONS

To effectively model and predict multivariate time series data such as that generated by coupled dynamical systems, it is important to determine the necessary inputs to the model and remove those inputs not required. Traditional methods of input variable selection may not deal with signals that become mixed and hence dependent. Hence in this paper, we show that the recently introduced ICA method, in conjunction with a proposed new method of input variable selection, is ideally suited to reducing the dimensionality of the input space and hece in the performance in prediction problems.

This test has not been proposed before due to the fact that the inputs are often dependent. In this way, it overcomes computationally expensive tests such as model-based methods and in addition, reduces the overall number of tests required from $2^n - 1$ to n for an n input model. ICA provides a powerful approach to the problem of input variable selection in complex model building.

ACKNOWLEDGEMENTS

The first author would like to thank Thomas Trappenberg and Sethu Vijayakumar for valuable discussions.

[2] Note that the cross-cumulants may be negative. We distinguish dependent from independent signals by means of the absolute magnitude of the cumulants.

6. REFERENCES

[1] B. V. Bonnlander and A. S. Weigend, "Selecting input variables using mutual information and nonparamteric density estimation," in *Proceedings of the 1994 International Symposium on Artificial Neural Networks (ISANN'94)*, Tainan, Taiwan, 1994, pp. 42–50.

[2] R. M. Neal, *Bayesian Learning for Neural Networks*, Ph.D. thesis, Department of Computer Science, University of Toronto, 1995.

[3] J. Moody and J. Utans, "Principled architecture selection for neural networks: Application to corporate bond rating prediction," in *Advances in Neural Information Processing Systems*, John E. Moody, Steve J. Hanson, and Richard P. Lippmann, Eds. 1992, vol. 4, pp. 683–690, Morgan Kaufmann Publishers, Inc.

[4] S. Makeig, A.J. Bell, T.-P. Jung, and T.J. Sejnowski, "Independent component analysis of electroencephalographic data," Technical report, Naval Health Research Center and The Salk Institute, 1995.

[5] S. Amari, A. Cichocki, and H.H. Yang, "A new learning algorithm for blind signal separation," in *Advances in Neural Information Processing Systems 8 (NIPS*95)*, G. Tesauro, D.S. Touretzky, and T.K. Leen, Eds., Cambridge, MA, 1996, The MIT Press, pp. 757–763.

[6] A.J. Bell and T.J. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, 1995.

[7] A. Cichocki and L. Moszczyński, "New learning algorithm for blind separation of sources," *Electronics Letters*, vol. 28, no. 21, pp. 1986–1987, October 8 1992.

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

[9] J.F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proc. F.*, vol. 140, no. 6, pp. 771–774, December 1993.

[10] C. Jutten and J. Herault, "Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, pp. 1–10, 1991.

[11] K.J. Pope and R.E. Bogner, "Blind signal separation. I: Linear, instantaneous combinations," *Digital Signal Processing*, vol. 6, pp. 5–16, 1996.

[12] B. Widrow and S.D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1985.

[13] A.D. Back, E. Wan, S. Lawrence, and A.C. Tsoi, "A unifying view of some training algorithms for multilayer perceptrons with FIR filter synapses," in *Proc. of the 1994 IEEE Workshop Neural Networks for Signal Processing 4 (NNSP94)*, J. Vlontzos, J. Hwang, and E. Wilson, Eds., New York, NY, 1994, IEEE Press, pp. 146–154.

[14] C.L Nikias and A.P Petropulu, *Higher-Order Spectra Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1993.

[15] A. Belouchrani, K. Abed Meraim, J.F. Cardoso, and É. Moulines, "A blind source separation technique based on second order statistics," *IEEE Trans. on S.P.*, vol. 45, no. 2, pp. 434–44, Feb. 1997.