New Approach Based on Independent Component Analysis for Dams Displacement Monitoring

THEODOR D. POPESCU
National Institute for Research and Development in Informatics
8-10 Maresal Averescu Avenue, 011455 Bucharest
ROMANIA
pope@u3.ici.ro

Abstract: - Independent Component Analysis (ICA) is an emerging field of fundamental research with a wide range of applications such as remote sensing, data communications, speech processing and medical diagnosis. It is motivated by practical scenarios that involve multisources and multisensors. The key objective of ICA is to retrieve the source signals without resorting to any a priori information about the source signals and the transmission channel. ICA using second-order statistics and high-order statistics based techniques and the corresponding algorithms will be presented to perform the blind separation of stationary or cyclostationary sources. In the last part of the paper, a case study with real data having as subject dams displacements monitoring will be presented.

Key-Words: - Independent component analysis, Blind source separation, Second-order statistics, High-order statistics, Large dams monitoring.

1 Independent Component Analysis

1.1 Problem Formulation

Independent Component Analysis (ICA) is a statistical and computational technique, that can be seen as an extension to Principal Component Analysis (PCA) and Factor Analysis (FA), [1]. ICA is a much more powerful technique, capable of finding the underlying factors or sources when these classic methods fail completely. The data analysed by ICA could originate from many different kinds of application fields, including digital images, economic indicators and psychometric measurements.

The simple ICA model assumes the existence of n independent signals $s_1(t), \ldots, s_n(t)$ and the observation of as many mixtures $x_1(t), \ldots, x_n(t)$, these mixtures being linear and instantaneous, i.e.

$$x_i(t) = \sum_{j=1}^{n} a_{ij} s_j(t) \qquad (1)$$

for each $i = 1, n$. This is compactly represented by the mixing equation

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \qquad (2)$$

where $\mathbf{s}(t) = [s_1(t), \ldots, s_n(t)]^T$ is an $n \times 1$ column vector collecting the source signals, vector $\mathbf{x}(t)$ similary collects the n observed signals and the square $n \times n$ "mixing matrix" $\mathbf{A}$ contains the mixture coefficients. The ICA problem consists in recovering the source vector $\mathbf{s}(t)$ using only the observed data $\mathbf{x}(t)$, the assumption of independence between the entries of the input vector $\mathbf{s}(t)$ and possible some a priori information about the probability distribution of the inputs. It can be formulated as the computation of an $n \times n$ "separating matrix" $\mathbf{B}$ whose output $\hat{\mathbf{s}}(t)$

$$\hat{\mathbf{s}}(t) = \mathbf{B}\mathbf{x}(t) \qquad (3)$$

is an estimate of the vector $\mathbf{s}(t)$ of the source signals (see Fig. 1).

ICA is closely related to the method Blind Source Separation (BSS) or blind signal separation. A "source" means here an original signal, i.e. independent component. "Blind" means that we no

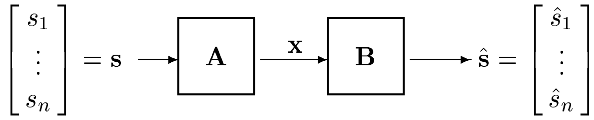

Fig. 1. Mixing and separating. Unobserved signals: $\mathbf{s}$; observations: $\mathbf{x}$; estimated source signals: $\hat{\mathbf{s}}$

very little, if anything, on the mixing matrix, and make little assumptions on the source signals. ICA is one method, perhaps the most widely used, for performing blind source separation.

In many applications, it would be more realistic to assume that there is some noise in the measurements, which would mean adding a noise term in the model:

$$\begin{aligned} \mathbf{y}(t) &= \mathbf{A}\mathbf{s}(t) \\ \mathbf{x}(t) &= \mathbf{y}(t) + \mathbf{n}(t) \end{aligned} \tag{4}$$

1.2 Identificability of the ICA model

The identificability of the noise-free ICA model has been treated in [2]. By imposing the following fundamental restrictions (in addition to the basic assumption of statistical independence), the identifiability of the model can be assured:

(1) All the independent components s_i with the possible exception of one component, must be non-Gaussian.
(2) The number of the observed linear mixtures m must be at least as large as the number of independent components n, i.e. $m \geq n$.
(3) The matrix $\mathbf{A}$ must be of full column rank.

For some algorithm classes these assumtions are not necessary. Usually, it is also assumed that $\mathbf{x}$ and $\mathbf{s}$ are centered. If $\mathbf{x}$ and $\mathbf{s}$ are interpreted as stochastic processes instead of simply random variables, additional restrictions are necessary. At the minimum, one has to assume that the stochastic processes are stationary in the strict sense. Some restriction of ergodicity with respect to the quantities estimated are also necessary.

In the ICA model of eq. (2), it is easy to see that the following ambiguities will hold:

(1) We cannot determine the variances (energies) of the independent components. The reason is that, both $\mathbf{s}$ and $\mathbf{A}$ being unknown, any scalar multiplier in one of the sources s_i could always be cancelled by dividing the corresponding column $\mathbf{a}_i$ in $\mathbf{A}$ by the same scalar. As a consequence we may quite as well fix the magnitudes of the independent components; as they are random variables, the most natural way to do this is to assume that each has unit variance: $E[s_i^2] = 1$. Then the matrix $\mathbf{A}$ will be adapted in the ICA solution methods to take into account this restriction.
(2) We cannot determine the order of the independent components. The reason is that, again both $\mathbf{s}$ and $\mathbf{A}$ being unknown, we can freely change the order of the terms in the sum (1), and call any of the independent components the first one.

1.3 Algorithms for ICA

Independent Component Analysis is mainly performed using the information on signal statistics. When the signals are temporal coherent, it is possible to solve the problem using only the second-order statistics. In this case the first assumption for the identificability of the ICA model, concerning Gaussian distribution of the sources, is not imposed.

If the signals are temporal white or have identical normalized spectral densities, without any information on a priori source distributions, the solution will need using of high-order statistics for the received signals. We underline that in the case of source signals temporal white and Gaussian, the blind source separation problem has not solution.

If the source signal distributions are known, the problem could be solved by maximum likelihood method. In this case, the second assumption for the identificability of the ICA model, concerning the number of independent components, is not imposed.

In the next two sections we present two approaches: the first supposes the signals temporal coherent and exploits the second-order statistics using intercovariance matrix of observations, and the second supposes the signals white temporal and exploits high-order statistics, using non-linear functions.

2 ICA Using Second-Order Statistics

2.1 Second-Order Statistics

The first step of the ICA procedure [3], consists of prewhitening the signal part $\mathbf{y}(t)$ of the observation. This is done via a whitening matrix $\mathbf{W}$, i.e. a $n \times m$ matrix (we consider n sources and m mixtures) such that $\mathbf{W}\mathbf{y}(t)$ is spatially white. The whiteness condition is

$$\mathbf{I}_n = \mathbf{W}\mathbf{R}_y\mathbf{W}^T = \mathbf{W}\mathbf{A}\mathbf{A}^T\mathbf{W}^T \tag{5}$$

where $\mathbf{I}_n$ denotes the $n \times n$ identity matrix. Equations (5) implies that $\mathbf{WA}$ is a unitary matrix: for any whitening matrix $\mathbf{W}$, it then exists a unitary matrix $\mathbf{U}$ such that $\mathbf{WA} = \mathbf{U}$. As a consequence, matrix $\mathbf{A}$ can be factored as

$$\mathbf{A} = \mathbf{W}^{\#}\mathbf{U} = \mathbf{W}^{\#}[\mathbf{u}_1, \ldots, \mathbf{u}_n] \tag{6}$$

where # denotes the pseudoinverse and $\mathbf{U}$ is unitary. The use of second-order information - in the form of an estimate of $\mathbf{R}_y$ which is used to solve for $\mathbf{W}$ in (5) - reduces the determination of the $m \times n$ mixing matrix $\mathbf{A}$ to the determination of a unitary $n \times n$ matrix $\mathbf{U}$. The whitened process $\mathbf{x}_w(t) = \mathbf{W}\mathbf{x}(t)$ still obeys a linear model:

$$\begin{aligned}\mathbf{x}_w(t) &\stackrel{def}{=} \mathbf{W}\mathbf{x}(t) = \mathbf{W}(\mathbf{A}\mathbf{s}(t) + \mathbf{n}(t)) = \\ &= \mathbf{U}\mathbf{s}(t) + \mathbf{W}\mathbf{n}(t)\end{aligned} \tag{7}$$

The signal part of the whitened process now is a unitary mixture of the source signals. Note that all the information contained in the covariance is 'exhausted' after the whitening, in the sense that changing $\mathbf{U}$ in (7) to any other unitary matrix leaves unchanged the covariance of $\mathbf{x}_w(t)$.

2.2 Whitening Matrix Computation

This step is implemented via eigendecomposition of the sample covariance matrix $\hat{\mathbf{R}}_x(0)$. We consider here that the noise covariance is of the form $\mathbf{R}_n(0) = \sigma^2\mathbf{I}_n$. The whitening procedure is the following:

(1) Estimate the covariance matrix $\hat{\mathbf{R}}_x(0)$ using T samples of the observations:

$$\hat{\mathbf{R}}_x(0) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}(t)\mathbf{x}(t)^T, \tag{8}$$

(2) Perform the eigendecomposition of the $\hat{\mathbf{R}}_x(0)$ covariance matrix

$$\hat{\mathbf{R}}_x(0) = \mathbf{H}\Delta\mathbf{H}^T \tag{9}$$

where

$$\mathbf{H} = [\mathbf{h}_1, ..., \mathbf{h}_m]$$

and

$$\Delta = diag[\lambda_1, \ldots, \lambda_m]$$

with $\lambda_i \geq \lambda_j$ for $i < j$. The number of sources can be estimated starting from the spectrum Δ, [4].

(3) Estimate noise variance $\hat{\sigma}^2$ as the average of the $m - n$ smallest eigenvalues of Δ

$$\hat{\sigma}^2 = \frac{1}{m-n}\sum_{i=n+1}^{m}\lambda_i \tag{10}$$

(4) Compute the whitening matrix $\hat{\mathbf{W}}$ as:

$$\hat{\mathbf{W}} = \Delta^{'}\mathbf{H}^{'T} \tag{11}$$

where

$$\Delta^{'} = diag[(\lambda_1 - \hat{\sigma}^2)^{-1/2}, \ldots, (\lambda_n - \hat{\sigma}^2)^{-1/2}]$$

and

$$\mathbf{H}^{'} = [\mathbf{h}_1, \ldots, \mathbf{h}_n]$$

This resulted matrix is used to obtain the whitened process

$$\hat{\mathbf{x}}_w(t) = \hat{\mathbf{W}}\mathbf{x}(t), \quad t = 1, \ldots, T \tag{12}$$

2.3 Intercovariance Matrix Estimation

Starting from the whitened process $\mathbf{x}_w(t)$, K intercovarince matrices of this process are computed:

$$\hat{\mathbf{R}}_w(k) = \frac{1}{T-k}\sum_{t=k+1}^{T}\mathbf{x}_w(t)\mathbf{x}_w(t-k)^T \tag{13}$$

where $1 \geq k \geq K$. The resulted matrices are of $n \times n$ dimension, and the computation effort does not depend of number of sensors, m. The value of K will be selected to realize a trade off between the statistic efficiency and computation effort. The value of the delays used in computation depends also on the length of the signal correlations. If we have a priori information on spectral density of sources, the value of K can be optimal chosen.

2.4 Joint Diagonalization

Let $\mathbf{R}_w = \{\mathbf{R}_w(k) | 1 \leq k \leq K\}$ be a set of K matrices with common size $n \times n$. A joint diagonalizer of the set $\mathbf{R}_w$ is defined as a unitary maximizer of the criterion

$$C(\mathbf{U}) \stackrel{def}{=} \sum_{k=1}^{K}|diag(\mathbf{U}^T\mathbf{R}_w(k)\mathbf{U})|^2 \tag{14}$$

where $|diag(\cdot)|$ is the norm of the vector build from the diagonal of the matrix argument. The problem is solved by a generalization of Jacobi technique [5], [6], [7].

2.5 Mixing Matrix and Source Signals Estimation

Let $\hat{\mathbf{U}} = [\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_n]$ be the unitary matrix resulted by joint diagonalization. If the objective of the blind identification is source separation, a brute estimation of these can be computed by:

$$\hat{\mathbf{s}}(t) = \hat{\mathbf{U}}^T \hat{\mathbf{x}}_w(t) \tag{15}$$

To estimate the mixing matrix need to inverse the effect of whitening, and the mixing matrix can be estimated by

$$\hat{\mathbf{A}} = \hat{\mathbf{W}}^{\#} \hat{\mathbf{U}} \tag{16}$$

To obtain at the output of the separator a maximum signal/noise ratio the source signals are estimated by

$$\hat{\mathbf{s}}(t) = \hat{\mathbf{A}}^T \hat{\mathbf{R}}_x(0)^{-1} \mathbf{x}(t) \tag{17}$$

2.6 The Algorithm

The general scheme of the SOBI algorithm (Second Order Blind Identification) can now be described by the following steps:

Step 1. Form the sample covariance $\hat{\mathbf{R}}_x(0)$ and compute the whitening matrix $\hat{\mathbf{W}}$

Step 2. Whitening the data provided by the sensors:

$$\hat{\mathbf{x}}_w(t) = \hat{\mathbf{W}}\mathbf{x}(t), \quad t = 1, \ldots, T$$

Step 3. Estimate K intercovariance matrices $\hat{\mathbf{R}}_w(k)$ of $\hat{\mathbf{x}}_w(t)$ for different time delay $k = 1, \ldots, K$

Step 4. Jointly diagonalize the set of intercovariace matrices in a base $\hat{\mathbf{U}} = [\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_n]$

Step 5. Estimate the mixing matrix with

$$\hat{\mathbf{A}} = \hat{\mathbf{W}}^{\#} \hat{\mathbf{U}}$$

Step 6. Estimate the source signals by

$$\hat{\mathbf{s}}(t) = \hat{\mathbf{A}}^{\#} \hat{\mathbf{R}}_x(0)^{-1} \mathbf{x}(t)$$

Note that at the second step of the algorithm the observation dimension is reduced to n, the source number. It results that the intercovariance matrices estimation is performed in a space of reduced dimension.

3 ICA Using High-Order Statistics

In the basic approach to solve ICA problem, the temporal structure of the received signals is in fact omitted and $\mathbf{s}(t)$ and $\mathbf{x}(t)$ are regarded as realizations of random vectors $\mathbf{s}$ and $\mathbf{x}$. We seek the solution of the form (3).

The problem for solving the separating matrix $\mathbf{B}$ is somewhat simplified if we consider only one of the source signals at a time. From equation (3) it follows:

$$\hat{s}_i = \mathbf{b}_i^T \mathbf{x} \tag{18}$$

with $\mathbf{b}_i^T$ the i-th row of $\mathbf{B}$.

The problem is further simplified by performing a prewhitening of the data $\mathbf{x}$: the observed vector $\mathbf{x}$ is firstly linearly transformed to another vector whose elements are mutually uncorrelated and all have unit variance. It can be shown that after this step, $\mathbf{B}$ will be an orthogonal matrix.

A recent review of various information theoretic contrast functions for solving $\mathbf{B}$, like mutual information, negentropy, maximum entropy, and infomax, as well as the maximum likelihood approach is given in [8].

As an example of contrast functions, consider the case of maximizing the kurtosis $E\{\hat{s}_i^4\} - 3[E\{\hat{s}_i^2\}]^2$ of the estimated signals $\hat{s}_i$. Because we assumed that the estimated signals have unit variance, this reduces to maximizing the fourth moment $E\{\hat{s}_i^4\}$. Its gradient with respect to $\mathbf{b}_i$ is $4E\{(\mathbf{b}_i^T\mathbf{x})^3\mathbf{x}\}$. In a gradient learning type rule, the row $\mathbf{b}_i^T$ of the separating $\mathbf{B}$ would be sought using a version of this gradient, in which the expectation is dropped and the gradient is computed separately for each input vector $\mathbf{x}$. In addition, a normalization term would be needed that keeps the norm of $\mathbf{b}_i$, equal to one - remember that the matrix $\mathbf{B}$ would be orthogonal due to the prewhitening of the data $\mathbf{x}$.

A much more efficient algorithm is the following fixed point iteration, [9]:

(1) Take a random initial vector $\mathbf{b}(0)$ of norm 1. Let $k = 1$.
(2) Let $\mathbf{b}(k) = E\{\mathbf{x}(\mathbf{b}(k-1)^T\mathbf{x})^3\} - 3\mathbf{b}(k-1)$. The expectation can be estimated using a large sample of $\mathbf{x}$ vectors.
(3) Divide $\mathbf{b}(k)$ by its norm.

(4) If $|\mathbf{b}(k)^T\mathbf{b}(k-1)|$ is not close enough to 1, let $k = k + 1$ and go back to step 2. Otherwise, output the vector $\mathbf{b}(k)$.

The final vector $\mathbf{b}(k)$ given by the algorithm equals the transpose of one of the rows of the (orthogonal) separating matrix $\mathbf{B}$.

To estimate n independent components, we run this algorithm n times. To ensure that we estimate each time a different independent component, we use the deflation algorithm that adds a simple orthogonalizing projection inside the loop. Recall that the rows of the separating matrix $\mathbf{B}$ are orthogonal because of the prewhitening. Thus we can estimate the independent components one by one by projecting the current solution $\mathbf{b}(k)$ on the space orthogonal to the rows of the separating matrix $\mathbf{B}$ previously found.

This algorithm, with the whitening and several extensions, is implemented in Matlab in the FastICA package which is a public domain package. A remarkable property of the FastICA algorithm is that a very small number of iterations seems to be enough to obtain the maximal accuracy allowed by the sample data. This is due to the cubic convergence of the algorithm.

Another algorithm intensively used in practice is JADE (Joint Approximate Diagonalization of Eigen-matrices), [10]. It is a typically batch algorithm using tensorial techniques as eigenmatrix decomposition. The algorithm is quite complicated, requiring sophisticated matrix manipulation.

4 Case Study - Dams Displacement Monitoring

One of the main objectives for dams displacements monitoring is to detect any abnormal behaviour alteration as early as possible. Any change in a dam response to some loads may be due to a structural deterioration culminating with the dam collapse. A change detected in real time can be decisive for the possible strengthening works.

Experience presented in [11], [12], [13] shows that the values of gross measurements recorded for dams point out the superposition of the following main three components: time, hydrostatic load, and temperature (see Fig. 2).

The time or irreversible component corresponds to the evolution in time of the dam behaviour. It can be amortized (strengthened) or amplified (de-

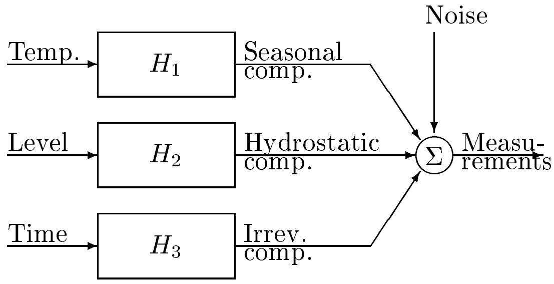

Fig. 2. Arch dam physical model

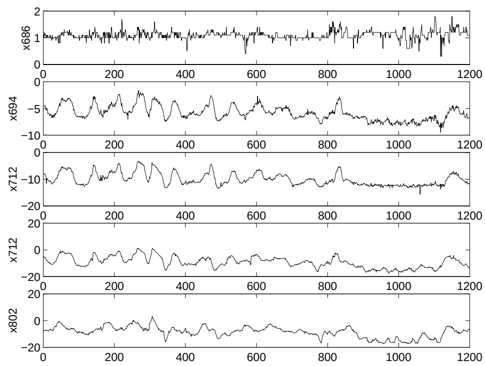

Fig. 3. Displacements for x axis at different levels

teriorated). The reversible hydrostatic component corresponds to the hydrostatic pressure effect of the lake level, while the reversible seasonal component depends on the distribution of temperatures and precipitation.

The objective of the application was to separate the components (sources) mentioned above starting from the displacements of the dam, without a priori knowledge of the generator phenomena or of the propagation environment, and by using only of the raw displacement measures. The application was dedicated to Vidraru dam, Romania, for a period of 1200 days.

The evolution of the dam displacements for x and y directions are given in Fig. 3 and Fig. 4, respectively, at different levels.

For these displacements, when SOBI algorithm has been used, resulted 3 independent sources which can be assimilated with the hydrostatic pressure component (lake level), seasonal component (temperature) and irreversible component. These are represented together with the lake level and temperature in Fig. 4. It can be noted that there are strong similarities between the estimated sources, representing seasonal and hydro-

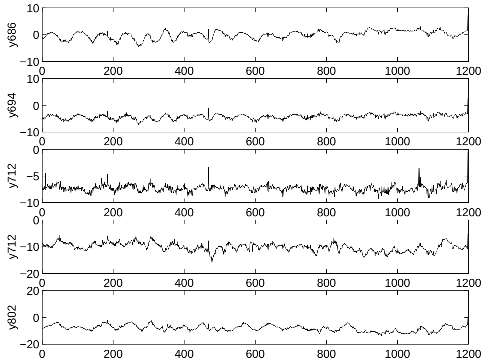

Fig. 4. Displacements for y axis at different levels

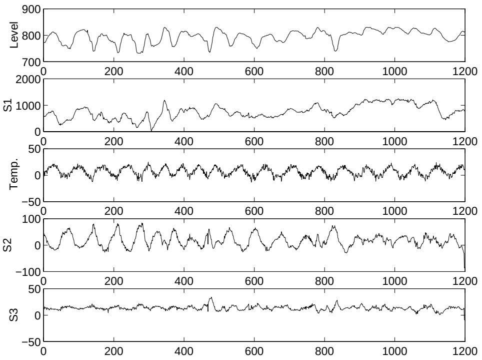

Fig. 5. Lake level, estimated source 1, temperature, estimated source 2 and source 3 assimilated with the irreversible component

static components, and temperature and lake level evolutions. The irreversible component, last represented, does not create special problems concerning dam safety.

The results represent only a preliminary analysis of the dam under study. More experiments and data analysis by different methods are necessary for a complete investigation of the dam behaviour.

5 Conclusions

The paper presented some methods and algorithms for independent component analysis based on second-order statistics and high-order statistics, to perform the blind separation of stationary or cyclostationary sources. The SOBI (Second Order Blind Identification) algorithm is described in detail and it is applied in an application having as subject displacements monitoring of an instrumented dam.

References:

[1] A. Hyvärinen, J. Karhunen and E. Oja, *Independent Component Analysis*, John Wiley & Sons, Inc., New York, 2001.

[2] P. Comon, "Independent Component analysis - a new concept ?", *Signal Processing*, vol. 36, pp. 287-314, 1994.

[3] 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*, vol. 45, pp. 434-444, 1997.

[4] Y. Yin and P. Krishnaiah, "Methods for detection of the number of signals", *IEEE Trans. on ASSP*, vol. 35, pp. 1533-1538, 1987.

[5] G. H. Golub and C.F.V. Loan (1989), *Matrix Computation*, The John Hopkins University Press, 1989.

[6] A. Souloumiac and J.F. Cardoso, "Comparaison de methodes de separation de sources", *Proc. GRETSI*, Juan les Pines, 1991.

[7] A. Souloumiac and J.F. Cardoso, "Givens angles for simultaneous diagonalization, *SIAM J. Matrix Anal. Appl.*, 1994.

[8] J. F. Cardoso, "Blind signal separation: statistical principles", *Proceedings of the IEEE*, vol. 9, pp. 2009-2025, 1998.

[9] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis", *Neural Computation*, vol. 9, pp. 1483-1492, 1997.

[10] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals, *IEE Proceedings - F*, vol. 140, pp. 362-370, 1993.

[11] P. Mazenot, *Methode generale d'interpretation des mesures de surveillance des barrages en exploitation a Electricite de France*, Division Technique Generale, 1971.

[12] D. Ispas, C. Scumpu, D. Hulea and Th. Popescu, "Dams and their foundations monitoring by statistic methods", *Hidrotehnica*, Special issue edited by The Romanian Committee on Large Dams, vol. 45, pp. 37 - 44, 2000.

[13] Th. Popescu, "Dams Displacements Monitoring Using Second Order Blind Identification Algorithm" , *Proc. IEEE International Symposium on Intelligent Control (ISIC)*, Vancouver, British Columbia, Canada, 27-30 October, 2002.