DELAY ESTIMATION IN SIGNAL SEPARATION OF TWO AND THREE DIMENSIONAL SIGNALS USING INDEPENDENT COMPONENT ANALYSIS

Vandana Singh1, Karan Singh2

J.K. Institute of Applied Physics & Technology

University of Allahabad (Allahabad), U.P., India

Email: vandana10.jk@gmail.com

SC&SS,Jawaharlal Nehru University, New Delhi, India

Email: karancs12@gmail.com

ABSTRACT

With the breakthrough of Independent Component Analysis (ICA) algorithm, it is possible now to even analyze the measured electromagnetic brain signals. Analysis with a conventional Independent Component Analysis (ICA) algorithm has previously demonstrated the results but the square mixing matrix assumption of conventional ICA causes large number of sources to be estimated. The modern ICA algorithm addresses this problem by extracting the signal sources which is closest in same sense, to a supplied reference signal. In this paper, we propose a method for delay estimation in separating speech of individual speakers from a multi-speaker speech signal using the knowledge of excitation source information. We examined and demonstrated our approach with two-microphone system for two and three dimensional signals. Through this paper, we present a technique for separating the audio sources from a single mixture. The system is based on the extraction of independent basis function from the mixture spectrogram and grouping them to produce the source subspaces. Principal component analysis is used for dimension reduction and independent component analysis is employed to make the basis functions independent from each other. The proposed algorithm is suitable for better grouping of the basis functions to separate the individual source. The satisfactory result of two-source mixture separation motivates to use our technique for real world single mixture source separation.

Keywords-*Independent component Analysis, delay estimation, analysis of signals, multi-speaker speech signal, novel approach*

1.INTRODUCTION

Models incorporating "latent" variables have been commonplace in the social and behavioral sciences for a long time. The most popular of those models is the factor analysis model, in which a set of observed continuous variables is explained in terms of a much smaller set of continuous latent variables (called factors), and the relationship is taken to be a linear one. Latent variables, which can be continuous or discrete, are quite different from observed variables in that they are artificial or hypothetical constructs.

Latent variables are typically used to give a formal representation of ideas or concepts that cannot be well-defined or measured directly. In educational and psychometric research, for example, fuzzy concepts such as "general intelligence," "verbal ability," "ambition," "socioeconomic status," "quality of life," and "happiness" is constructed from certain observed variables that are regarded as proxies for those unobservable concepts. Moreover, it is not unusual to hear of a causal relationship between a latent variable and a set of given observable variables (e.g., "it is because of a person's high level of intelligence that he or she does so well on standardized tests"). Latent variables are also known, for example, as hidden variables in neural network modeling and as sources that are statistically independent of each other in independent component analysis. Latent variables have been introduced into MCMC sampling as auxiliary variables and as a data-augmentation technique in missing-value problems. Latent variables are usually formed as linear combinations of observable variables for the purpose of reducing the dimensionality of a data set. Indeed, it is easier to consider a single latent variable interpreted as "quantitative ability" than to have to deal with understanding a battery of different arithmetic and mathematics test scores. As we will see, latent variables play the fundamental role of "sources" in blind source separation problems.

2. BLIND SOURCE SEPARATION

Used in sound and image processing, brain imaging, remote sensing, signal processing, stuck-market movement. An example of Blind source separation is cocktail party problem 'm' persons are speaking in a party simultaneously 'r' microphones placed at different distances in the same room and record a different mixture of speakers voice at different time points.

The problem is to separate out speech signals of each of each of 'm' speakers based upon these microphone recordings. Thus the problem is immixing the mixture of signals. Special cases of this model include independent component analysis.

3. INDEPENDENT COMPONENT ANALYSIS (ICA)

Independent component analysis (ICA) is a multivariate statistical technique that seeks to uncover hidden variables in high-dimensional data. As such, it belongs to the class of latent variable models. Furthermore, because of its success in analyzing signal processing data, ICA is also regarded as a digital signal transform method. In its most basic form, the ICA model is assumed to be a linear mixture of an unknown number of unknown hidden source variables, where the mixing coefficients are also unknown. A totally “blind” approach to determining both the hidden variables and the mixing coefficients solely from the observed multivariate data fails because the problem as stated is not well-defined.

To build more structure into the problem, we require the hidden variables to be mutually independent and also (with at most one exception) non-Gaussian. ICA is actually an amalgam of several related approaches to this problem, and these approaches are characterized by the types of assumptions visited upon the distributions of the independent source variables and whether or not a separate noise component should be included in the ICA model.

3.1 ICA Model

	APPLICATIONS
EEG	Relate certain type of behavior of changes in electrical action / activity of cerebral cortex.
ERP	Fine tuned EEG, resulting from simulation of visual, auditory or sensory systems.
MEG	Strength of magnetic field generated by cortical activity
FMRI	(Functional Magnetic Resonance) used to study human brain, web image, retrieval & classification, remote sensing, classification of micro array, gene profile.

Table 1. ICA general and medical applications

Linear mixture of unknown number of unknown variables, where mixing coefficients are also unknown, hidden variables are assumed to be non Gaussian and independent. ICA has vast no. of applications. We come across many medical applications like EEG, etc. in our life. It has made the estimation of signals particularly in medical diagnosis very efficient.

3.1 General ICA problem

In general form we assume that $\tilde{X}$ is generated by

$\tilde{X}=g(\tilde{s})+\tilde{e}$

Where $\tilde{S}$ = S1 : mX1 vector of unobservable source

. component (latent source variables)

.

Sm

$\tilde{X}$ = X1 :rX1 vector of observation

.

.

Xr

g: mixing function

$\tilde{e}$ = e1

. = error

.

er

g : RmRr

BSS problem is to invert g & find $\tilde{s}$

Sj = jth latent variable :ammeel to have mean 'zero' & variance '1'

If $g(\tilde{s}) = A\tilde{s}$ where 'A' in a r X m matrix of mixing parameter then model is linear ICA model.

If $g(\tilde{s})$ is non-linear we have non-linear ICA model.

* (mixing function is unchanged with time)

If 'e' is not included then all noise is associated with s and model is called noiseless ICA if 'e' is included, then model is called "noisy ICA".

Let us consider ICA with linear structure

$$\tilde{X} = A\tilde{s}$$

We assumed that s is non-Gaussion and has independent components. For Gaussian case, it is not possible to estimate A & $\tilde{S}$ separately.

Xi = Xi1

. i = 1,2,….n : dataset

.

Xir

our is to recover 'm' independent sources.

Si = Si1

. ; i = 1,2,…. n

Sim

(linearly independent)

and $\tilde{X}i = Asi$

A = rXm

usually msr & rank of A=m

If 's' has mean 'o' and covariance matrix 'Im'.Then $\tilde{X}$ has mean vector 'o' &covariance matrix AA'

premultiplying both side of

$$\tilde{X} = A\tilde{S}$$

By $(A'A)A$ we obtain

$$(A'A)A'\tilde{X} = A\tilde{S}(A'A)A'$$

$$= \tilde{S}$$

or $\tilde{S} = W\tilde{X}$

where

w = (A'A)A':unmixing/separating matrix

Note: if A is square matrix = (A'A)A'

=A(A')A'

=A

Then

$$Sk = Wk\tilde{X} \quad (k=1,2\ldots\ldots,m)$$

W'k is kth row of W

If $\hat{}$w is an estimate of 'w' then we can estimate s by

S=wx

4. The FAST ICA Algorithm

1) Centre & whiten the data.

2) choose an initial vector 'w' with unit norm

$(\|w\| = (\tilde{w}'\tilde{w}) = 1)$

$\tilde{w}$ may be chosen randomly

3) Choose G to be any non-quadratic density with first

& second partial derivatives 'g' & 'g''

(some choices are log cosh & exp)

	G(y)	g(y)	g'(y)
log cosh	$1/\alpha$ log cosh(αy)	tanh(αy)	α(1 - tanh2(αy))
exp	- *e*- *y*2/2	*ye*- *y*2/2	(1-*y*2)*e* - *y*2/2

Table 2. the log cosh and exponential values of G(y), g(y), and g'(y)

4) (Modify the weight in average form)

5) w = ||w||

6) Return to steps 4&5 until convergence is achieved

FAST ICA package uses two different ways for extracting components.

4.1 Deflation

The single component routine finds a new component that new component is orthogonalized with respect to all previously found compounds and then resulting component is normalized.

4.2 Parallel

The single component routine is carried out in parallel for each independent component to be extracted and then a symmetric orthogonalization is carried out and all components simultaneously.

4.3Deflation Algorithm

1) Centre & whiten data to give X

2) Decide a no. of components 'm' to be extracted.

3) For K=1,2,…..,m

-initialize r-vector(r*1) wk with ||wk||=1

- let

wk

(FAST ICA single compound update)

- let wk

- iterate wk until convergence achieved

4) Set K K+1 if ksm returns to step 3

4.4 Parallel algorithm

1) Centre & whiten data to give X

2) decide (no. of component) 'm'.

3) Initialize r vector w1,w2,…..wm each ||wk|| = 1

4) Carry out symmetric orthogonalization .

5) For K=1,….m

(weights update)

6) Carry out symmetric orthogonalization for 'w'.

7) Return to step 5 until convergence is achieved.

5. EXPERIMENT

For finding out the quality of recovery of signal, we performed various experiments which are based on three parameters which are:

(1) Number of signal

(2) Number of method

(3) Number of dimensionality

Now we will take these three parameters to check which signals recovery is best and whose recovery is worst.

5.1 TWO DIMENSIONAL SIGNAL

First we generated signals of two dimensional in which we have selected height and width randomly. In two dimensional we use images to watch the figure.

The original signal which we ran on FASTICA, following things got separated

(a) original signal

(b) Mix signal

(c) Recovering signal

Now we have to find out quality of recovery of signal which is based on three parameters which are:

(1) Number of signal

(2) Number of method

(3) Number of dimensionality

Now we will take these three parameters to check which signals recovery is best and whose recovery is worst.

Suppose: we took number of signal two. To check it we took six numbers of methods. In that we find out same get very good recovery and some get bad recovery. After six numbers of methods we find out Q (sum of square root of original signal and recovery signal) of all the methods separately.

(1) Then we find out mean individually

(2) Then finally the mean of all the means is calculated.

Similarly with the method written below, we check it till the five signals. Then we find out that as we are increasing the number of signals, the mean of Q is also increasing and the standard derivation is decreasing.

5.1.1 OUTPUT OF TWO SIGNALS

S. no	Original Signal	Recovery Signal	value	Individual Mean	Std.
1.	1 2	2 (+) 1 (-)	0.0158 0.0028	Q=0.0093	0.0093
2.	1 2	1 (+) 2 (+)	0.0142 0.2700	Q=0.1421	0.1809
3.	1 2	2 (-) 1 (+)	0.0011 0.0094	Q=0.0053	0.0059
4.	1 2	1 (+) 2 (-)	0.0147 0.0070	Q=0.0106	0.0054
5.	1 2	1 (-) 2 (+)	0.1624 0.0896	Q=0.1260	0.0515
6.	1 2	2 (+) 1 (-)	0.0023 0.0245	Q=0.0134	0.0157

Table 3. The statistical data obtained from output of two signals in 2-D signals

S. No	Original Signal	Recovery Signal	Value	Individual Mean	Std.
1.	1 2 3	3 (-) 1 (+) 2 (+)	0.0232 0.0111 0.2399	Q=0.0914	0.1287
2.	1 2 3	2 (+) 1 (+) 3 (+)	0.0350 0.0535 0.0268	Q=0.0384	0.0137
3.	1 2 3	1 (+) 2 (+) 3 (-)	0.0511 0.1312 0.0934	Q=0.0919	0.0401
4.	1 2 3	1 (+) 2(+) 3 (-)	0.1605 0.0056 0.0189	Q=0.0617	0.0859

Table 4.
The statistical data obtained from output of three signals in 2-D signals

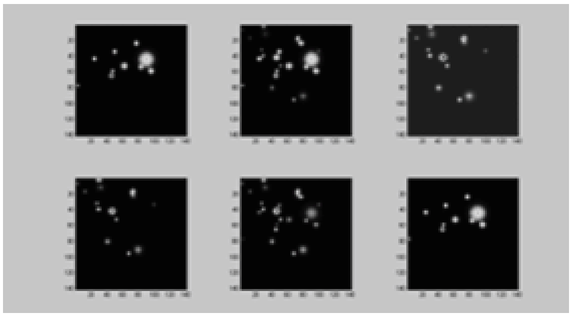

Fig1. Matlab experimentation view of output of three signals of 2-D signal

5.1.2 OUTPUT OF THREE SIGNALS

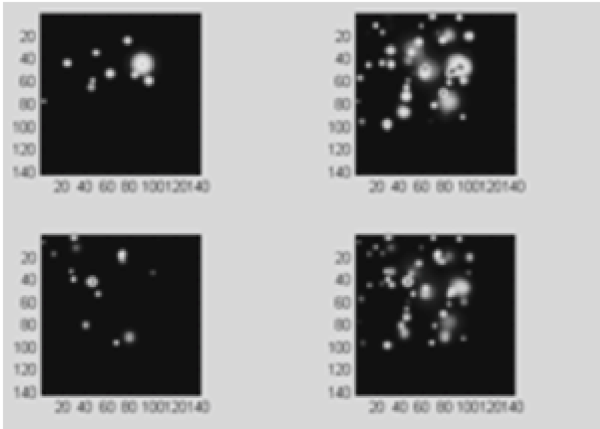

Fig2.
Matlab experimentation view of output of two signals of 2-D signal

5.1.3 OUTPUT OF FOUR SIGNALS

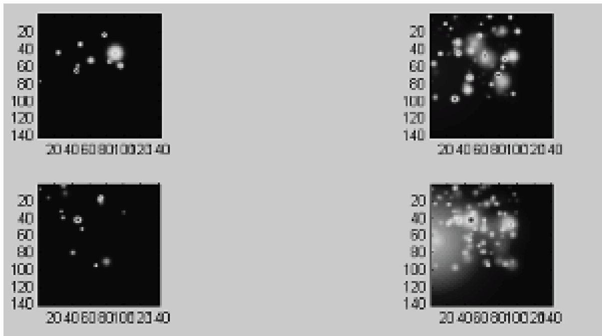

Fig3.
Matlab experimentation view of output of four signals of 2-D signal

S. No.	Original Signal	Recovery Signal	Value	Individual Mean	Std.
1.	1 2 3 4	2 (-) 1 (-) 3 (+) 4 (-)	0.0955 0.0123 0.2032 0.1884	Q=0.1259	0.885
2.	1 2 3 4	2 (+) 1 (+) 4 (+) 3 (+)	0.0666 0.0333 0.1393 0.0317	Q=0.0677	0.0504
3.	1 2 3 4	1 (-) 2 (-) 3 (-) 4 (-)	0.0135 0.1353 0.2261 0.0615	Q=0.0787	0.0927
4.	1 2 3 4	1 (-) 4 (-) 2 (-) 3 (-)	0.0743 0.1964 0.1819 0.1115	Q=0.1410	0.0579

Table 5. The statistical data obtained from output of four signals in 2-D signals

5.2 THREE DIMENSIONAL SIGNAL

Three dimensional is similar to the two dimensional. But in this till two dimension features are clear but as we increase the number of dimensionality are not clear. Therefore, we convert three dimensional into two dimensional. We use slice to watch the figure but as we increase the number of signals so features are not clear.

5.2.1 OUTPUT FOR TWO SIGNALS

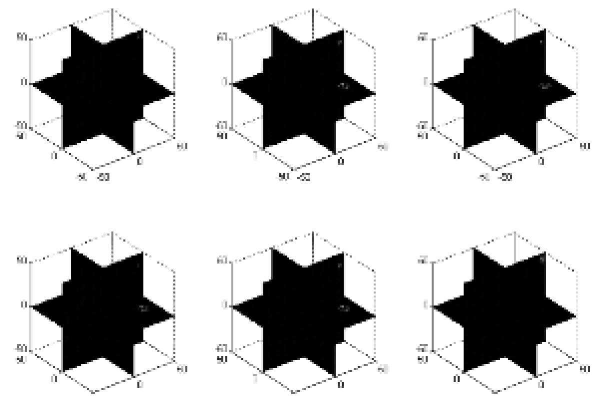

Fig 4.
Matlab experimentation view of output of two signals of 3-D signal

Then increase the number of signal features are not clear. Therefore we convert three dimensional into two, dimensional. To do this we take projection.

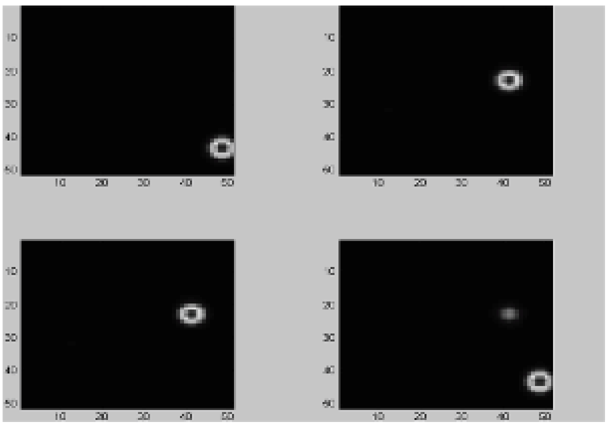

Fig5.
Matlab experimentation view of output of two signals of 3-D signal

5.2.2 OUTPUT FOR TWO SIGNALS PROJECTION

5.2.3 OUPUT OF THREE SIGNALS

S. No.	Original signal	Recovery Signal	Value	Individual Mean	Std.
1	1 2	2 (-) 1 (+)	0.0034 0.0031	Q=0.0033	0.0389
2	1 2	1 (+) 2 (+)	0.0024 0.0026	Q=0.0025	0.0259
3	1 2	2 (+) 1 (-)	0.0050 0.0013	Q=0.0032	0.0026
4	1 2	1 (-) 2 (+)	0.0013 0.0012	Q=0.0013	0.0476

Table 6.
The statistical data obtained from output of two signals in 3-D signals

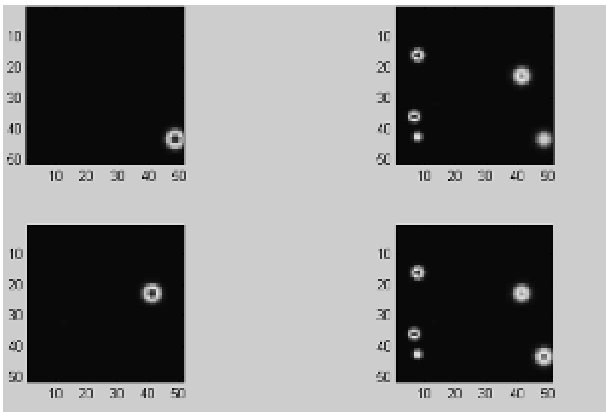

Fig6.
Matlab experimentation view of output of three signals of 3-D signal

S. No.	Original signal	Recovery Signal	Value	Individual Mean	Std.
1	1 2 3	1 (-) 2 (-) 3 (+)	0.0124 0.0572 0.0860	Q=0.0519	0.0371
2	1 2 3	3 (+) 2 (+) 1 (+)	0.0089 0.0114 0.00011	Q=0.0071	0.0054
3	1 2 3	1 (+) 2 (+) 3 (-)	0.0116 0.0480 0.0091	Q=0.0229	0.0218
4	1 2 3	3 (-) 1 (+) 2 (-)	0.0301 0.0136 0.0170	Q=0.0202	0.0087

Table 7. The statistical data obtained from output of three signals in 3-D signals

5.2.4 OUTPUT FOR FOUR SIGNALS – PROJECTION

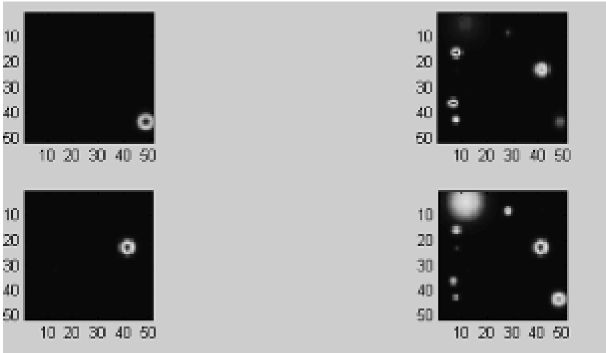

Fig7. Matlab experimentation view of output of four signals of 3-D signal

S. No.	Original signal	Recovery Signal	Value	Individual Mean	Std.
1	1 2 3 4	2 (+) 1 (-) 3 (+) 4 (-)	0.0146 0.0582 0.0912 0.0544	Q=0.0546	0.0134
2	1 2 3 4	1 (-) 2 (+) 4 (-) 3 (-)	0.0809 0.0141 0.0091 0.0852	Q=0.0473	0.0413
3	1 2 3 4	1 (-) 2 (-) 3 (-) 4 (+)	0.0210 0.0293 0.0291 0.0151	Q=0.0236	0.0069
4	1 2 3 4	2 (-) 3 (+) 4 (+) 1 (-)	0.0150 0.0395 0.1048 0.0149	Q=0.0200	0.0424

Table8. The statistical data obtained from output of four signals in 3-D signals

6. RESULT COMPARISON:

From our experimentation of fast ILA over two microphone system for determining the delay estimation of multiple signals in 2-D and 3-D, we infer that the following result proofs our project and tell us that our process can be set as an example for further growth in ICA related estimations. The result of all the statistical data obtained from our experimentation of 2-D and 3-D signals in given in table 9.

	Two Dimensional signals		Three Dimension signals	
No. of Signals	Mean	Std. Deviation	Mean	Std. Deviation
Q2	0.0511	0.0691	0.0035	0.0211
Q3	0.0826	0.0459	0.0196	0.0137
Q4	0.0997	0.0266	0.0282	0.0185

Table 9. Comparative statistical of 2-D and 3-D signals of two-microphone system in MATLAB

7. CONCLUSION

Independent component analysis (ICA) is a multivariate statistical technique that seeks to uncover hidden variables in high-dimensional data. As shown in the project that we have used Fast ICA technique for signal separation which is of two types deflection and parallel. Further in Fast ICA, we have used deflection technique.

Signals which is being used here is Gaussian signal. As we know that in non-Gaussian signals, there is a perfect overlapping of signals which makes its separation difficult. Therefore Gaussian signal is used for signal separation.

Thus ICA is used to reduce the dimension of signal. It is a very useful technique for signal separation which makes it extremely useful in medical field.

REFERENCE

1. Beckmann, C.F., Smith, S.M.,. Tensorial extensions of independent component analysis for group fMRI data analysis. NeuroImage, Vol. 25. No. 1, pp 294-311, 2005.
2. Calhoun, V.D., Liu, J., Adali, T.,. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. NeuroImage, Vol. 45. No. 1, pp 263, 2009.
3. Himberg, J., Hyvärinen, A., Esposito, F.,. Validating the independent components of neuroimaging time-

series via clustering and visualization. NeuroImage, Vol. 22. No. 3, pp 1214–1222, 2004.

4. Hyvärinen, A., Ramkumar, P., Parkkonen, L., Hari, R.,. Independent component analysis of short-time Fourier transforms for spontaneous EEG/MEG analysis. NeuroImage, Vol. 49, No. 1, pp 257-271, 2010.

5. van de Ven, V.G., Formisano, E., Prvulovic, D., Roeder, C.H., Linden, D.E.,. Functional connectivity as revealed by spatial independent component analysis of fMRI measurements during rest. Human Brain Mapping, Vol. 22. No. 3, pp 165–178, 2004.

6. Hyvanrinen,. Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural Network, Vol. 10. No. 3, pp 626–634, 1999.

7. Hyvanrinen, E. Oja,. Independent component analysis: algorithms and applications. Neural Network, Vol. 13. No. 5, pp 411–430, 2000.

8. Independent component analysis: theory and applications. Kluwer Academic Publishers, 1998.

9. E. H. Wu and P. L. Yu, “Independent component analysis for clustering multivariate time series data,” Vol. 35, No. 6, pp 474–482, 2005

10. S. Fiori, “Overview of independent component analysis technique with an application to synthetic aperture radar (sar) imagery processing”, Neural Network, vol. 16, no. 3-4, pp. 453–467, 2003.

11. Groppe, D.M., Makeig, S., Kutas, M.,. Identifying reliable independent components via split-half comparisons. NeuroImage, Vol. 45. No. 4, pp 1199–1211, 2009.

12. P. Comon, “Independent component analysis—a new concept?”, Signal Processing, Vol. 36, pp 287–314, 1994.

13. T.-W. Lee, M. Girolami, T. J. Sejnowski, , “Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and super-gaussian sources”, Neural Computation , Vol. 11, No. 2, pp 417-441, 1999.

14. A. Hyvarinen, J. Karhunen and E. Oja, "Independent Component Analysis", Wiley & Sons, 2001.