USING ICA FOR ANALYSIS OF SEISMIC EVENTS

Ramaswany Thiyagu, Andrew Chanerley
School of Computing, Information Technology and Engineering
University of East London

Abstract: Independent Component Analysis (ICA) is a statistical and computational technique for revealing hidden factors that underlie set of random variables, measurements, or signals. ICA is a general purpose technique which is used to linearly transform the observed random data into components. The ICA can be estimated by using the concept of maximum non-Gaussianity, maximum likelihood estimation, or minimisation of mutual information. This paper applies ICA to seismic acceleration time histories in order to locate any hidden components of ground rotational motion or tilts. Normally the three components of seismically induced rotations are not recorded in most of the available seismic instruments, primarily because previous devices did not provide the required sensitivity to observe rotations in a wide frequency band and distance range (the two horizontal components, equal to tilt at the free surface, are generally recorded at low frequencies) Igel *et al* 2003. From the x, y and z components usually recorded the Extended Generalised Lambda Distributions (EGLD) – ICA model was used to examine whether rotational or tilt trends were embedded within the 3 components. The algorithm tries to fit a matrix from the data which will separate any other trends within the available components. The results show that the EGLD-ICA separates trends within the 3 components; however these are not yet identified as tilts or rotations.

1. Introduction:

Basic seismic data for earthquake engineering is obtained from measurements of tremor during earthquakes. The seismic data are sampled over the duration of earthquake and stored in the equipment (Seismograph). The data is inevitably smeared with background noise with both the short and long frequency range. The first ever attempt to develop a procedure to correct recorded Seismograph was made in the 1970's by Trifunac *et al*. The earthquake records from U.S. events are widely available but not the European records so the European council, Environment and Climate Research Programme have disseminated a database of corrected European records (Ambraseys *et al*. 2000). There is a particular correction technique, i.e. band-pass filtering of the data which is used to remove noise outside the ground motion of the band of interest The wavelet transform is also used (Daubechies 1992) to de-noise the time history (Donoho 1992) as an alternative for band-pass filtering. The de-noising removes low and high frequency corrupting signals but retains relevant data giving a better estimation of true ground motion.

In engineering applications, there are many distributions of importance that are non Gaussian and skewed. Furthermore, these non-Gaussian distribution may have a Gaussian fourth moment i.e. kurtosis near zero. Fixed-point algorithm EGLD-ICA model is proposed for possible separation of Seismic unknown components. The separated signals are yet to identify as tilts or rotation. Simulation results shows the reliable performance of an EGLD-ICA based algorithm in separating sources with commonly used distribution in signal processing.

This paper gives a brief description of ICA, Fast-ICA and BSS models in sections 2, 3 and 4 respectively. In section 5, the EGLD-ICA model is used in separating the signals from the original source in particular the output from a three degree-of-freedom seismograph which is used in section 6. In section 7 the simulation results of the separated signals are shown.

1.1. Problem Description:

Blind Source Separation (BSS) by Independent Component Analysis (ICA) has received attention because of its potential application in signal processing, medicine and economics. The corrected seismic data will be used in the BSS processor (Ref: fig 4.1) to find a mixing and de-mixing matrix with non-Gaussian statistics in different configuration with different approaches. The output data from the model will be compared with Fast-ICA for further application. Principal Component Analysis (PCA) is the pre-cursor to ICA where transformed variables are as statistics independent from each other as possible. Thereby providing alternative data to analysis and find any other trends within the available components.

2. Independent Component Analysis:

ICA is a statistical and computational technique for revealing hidden factors that underlie set of random variables, measurements, or signals. It is a method for finding underlying factors or components from multidimensional statistical data (Hyvarinen 2000).
ICA is an extension of linear transform called Principal Component Analysis (PCA). PCA was developed some years ago in context with Blind Source Separation (BSS) in Digital Signal Processing (DSP) and array processing. ICA is solved on the basis of optimization of certain measures of departure from Gaussanity, which leads to a numerical optimization problem. ICA imposes statistical independence on the individual components of the output vector and has no orthogonality constraint (Mutihac 2003). Furthermore Blind Source Separation algorithm from ICA, which operates on data with non-Gaussian statistics, will then de-mix the mixed seismic data and thereby providing the unknown component. These EGLD-ICA (Extended Generalized Lambda Distribution) data is compared with Fast-ICA for further analysis.

3. Fast-ICA:

Fast ICA is an efficient and popular algorithm for independent component analysis invented by Hyvärinen *et al* 2000 at Helsinki University of Technology. The algorithm is based on a fixed-point iteration scheme maximizing non-Gaussanity as a measure of statistical independence.
In a linear BSS model, we observe ***m*** signals $x_1, x_2, ..., x_m$ that correspond to a linear mixture of a ***p*** source signal
$s_1, s_2, ..., s_p$, i.e.

$$x = As, \qquad (1)$$

where A is known as the ***mxp*** ''mixing matrix''. According to (eq: 1) and given the observable vector **x**, a linear projection is performed as

$$y = w^t x, \qquad (2)$$

where, clearly, ***w*** must tend towards one of the column vectors of A^{-1} in order to obtain one of the ***p*** source signals s_i with $i \in \{1, ..., p\}$. If there is a minimum of

$(p-1)$ non-Gaussian sources and A is forced to be orthonormal by a previously performed whitening process, the so-called fixed-point or fast-ICA algorithm recovers one of the original signals (Bermejo 2007). The EGLD-ICA is compared to Fast-ICA algorithm with different contrast functions with the symmetric approach. The deflation approach of the Fast-ICA seems to give similar results.

4. Blind Source Separation:

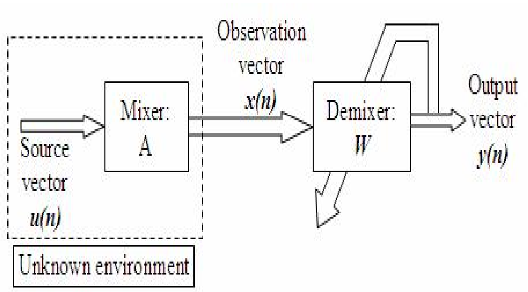

Fig 4.1 Processor for the BSS

It is assumed that the source signals $U_1, U_2, ..., U_m$ are zero-mean signals, which in turn means that the observables $X_1, X_2, ..., X_m$ are also zero-mean signals. The same result is true for the de-mixer outputs $Y_1, Y_2, ..., Y_m$ (Haykin 1999). The information used to recover the original signal sources is contained in a realization of the observation vector **X**, denoted by **x**. the underlying principle involved in its solution is called Independent component analysis (ICA) (Comon, 1994), which may be viewed as an extension of principal components analysis (PCA). ICA imposes statistical independence on the individual components of the output vector **Y** and has no orthogonality constraint. It separates mixed signals into independent signals, which are non-linearly uncorrelated. ICA algorithms estimate how Gaussian distributed the joint distribution of the independent components is. The less Gaussian the distribution, the more independent the individual components actually are. To perform the task of source separation in an unsupervised manner, we use by two methods, maximum likelihood and maximum entropy.

4.1. Maximum Likelihood Method:

Maximum Likelihood Estimation method is a technique to solve the blind source separation problem. It as a well established procedure for statistical estimation. This procedure works by formulating the log-likelihood function and then optimizing it with respect to the parameter vector of the probabilistic model.

4.2. Maximum Entropy Method:

The maximum entropy method for blind source separation is due to Bell and Sejnowski 1995, the system based on this method. As before, the de-mixer operates on the observation vector **X** to produce an output **Y** = **WX** that is an estimate of the original source **U**.

5. Extended Generalized Lambda Distribution (EGLD):

The EGLD algorithms employing the maximum likelihood principle and can separate a relatively wider class of non-Gaussian source signals, even skewed distributions with zero kurtosis. EGLD algorithms are computationally fast, as both the estimators for parameters and the score function are simple rational functions (Karhunen and Koivunen 2002). Signals with kurtosis equal to that of the Gaussian distribution can be separated by using the EGLD-ICA algorithm.

The Extended Generalized Lambda Distribution (EGLD) model is used for modelling the source distributions. It is a large family of distributions covering the whole space of the third and the fourth moments. In 1960, Turkey presented the lambda distribution concept which was generalized in 1970’s. The latest extension is done by Karian *et al.* in 1996 which is a combination of Generalized Lambda Distribution (GLD) and Generalized Beta Distribution (GBD). EGLD covers the most important distributions including normal, uniform, gamma and beta distributions.
Finally, ICA using an EGLD model will find underlying source distributions, estimated through the marginal distributions that are fitted into the EGLD family. This process is being repeats until convergence (Karian et al. 1996).

6. Seismograph (3 Degree of freedom):

At present, there are two types of measurements that are routinely used to monitor global and regional seismic wave fields. First, standard inertial seismometers measure three components of translational ground displacement (velocity, acceleration) and form the basis for monitoring seismic activity and ground motion. The second type aims at measuring the deformation of the Earth (strains). It has been noted for decades (Aki and Richards, 1980, 2002) that there is a third type of measurement that is needed in seismology and geodesy in order to fully describe the motion at a given point, namely the measurement of ground rotation (a vectorial quantity). Specifically, if u(x) is the displacement at position x, the displacement at an arbitrarily close position is given by

$$u(x+\delta x)=u(x)+D\delta x=u(x)+\in\delta x+\omega*\delta x,$$

where D is the deformation gradient, e is the - symmetric - strain matrix and $\omega=(1/2)\Delta*u$ is the rotation (also sometimes called spin or vorticity). The three components of seismically induced rotation have been extremely difficult to measure, primarily because previous devices did not provide the required sensitivity to observe rotations in a wide frequency band and distance range (the two horizontal components, equal to tilt at the free surface, are generally recorded at low frequencies). Indeed, (Aki and Richards 2002) note that ‘‘seismology still awaits a suitable instrument for making such measurements’’ according to Igel *et al* 2003.

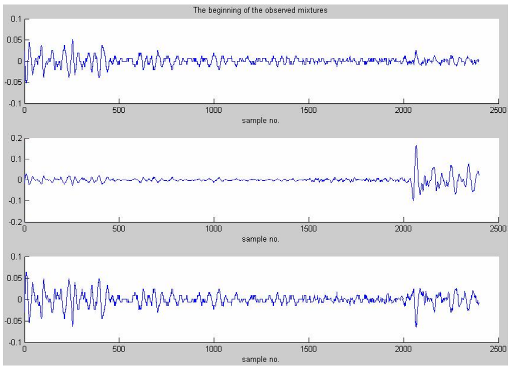

Fig 6.1 Seismic data from the problematic device

From the figure 6.1, shows the seismic data corrupted with noise due to lack of sensitivity of 3 degree-of-freedom instrument; therefore it’s difficult to separate the unknown component to identify tilts or rotations.

7. Simulation Results:

The performance of the EGLD-ICA algorithm is as follows: we consider first an example with a mixture of three sources, a sine wave (sub-Gaussian), Seismic data signal (super-Gaussian), and a random Gaussian with zero mean and unit variance. The sample size is 7200, and the random matrix

$$A = \begin{pmatrix} 0.1746 & -0.5883 & 0.1139 \\ -0.1867 & 2.1832 & 1.0668 \\ 0.7258 & -0.1364 & 0.0593 \end{pmatrix}. \quad (7.1)$$

The simulation results are shown in fig 7.1, 7.2, 7.3. It can be seen that the sources are well separated.

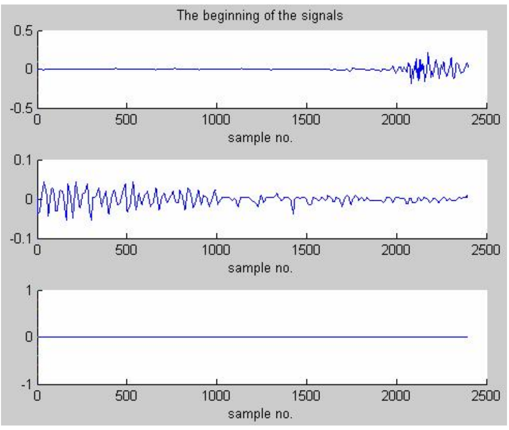

Fig 7.1 Source Signal

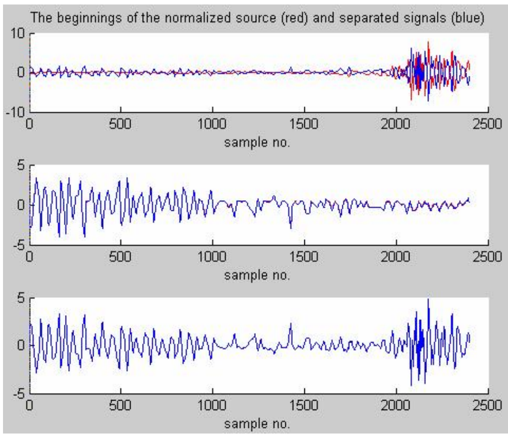

Fig 7.2 Mixed Signals

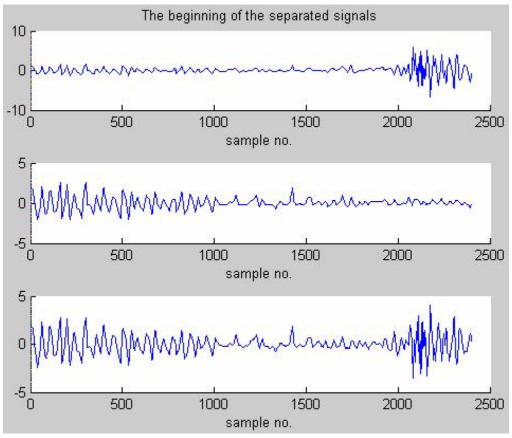

Fig 7.3 Separation by the EGLD-ICA

The EGLD-ICA algorithm can separate signals with kurtosis equal to that of Gaussian distribution. To illustrate this, three signals of sample length 7200 is generated. The sources are mixed using the randomly generated matrix (7.1). From the fig 7.3 the EGLD-ICA reliably separates all the sources.

8. Conclusions and Future Work:

The EGLD-ICA algorithm for ICA estimates the independent component from given multidimensional Seismic data signals. Based on the Blind Source Separation method of Independent Component Analysis is able to identify the Mixing and De-mixing matrix with non-Gaussian statistics in different configuration for particular soil characteristics. The Maximum Likelihood method from Haykin (1999) is able to separate more general range of source distribution, which allows an extension to the separation of mixtures with non-Gaussian sources and used Principal Component Analysis for linear transform to get transformed variables as statistically independent from each other as possible. The transformed non-linear data is provided for further analysis in ICA. The

Fast-ICA algorithm seems to give similar results and tries to fit a matrix from the data which will separate any other trends within the available components. The result shows that the EGLD-ICA separates trends within the 3 components; however these are not yet identified as tilts or rotations. Future analysis could be done on Seismic data collected from higher degree-of-freedom measuring instruments.

9. Reference:

Aki, K., and P. G. Richards (1980), Quantitative Seismology, 1st Edition, W. H. Freeman, San Francisco, California.

Aki, K., and P. G. Richards (2002), Quantitative Seismology, 2nd Edition, University Science Books, Sausalito, California.

Alexander N.A., Chanerley A.A., and Goorvadoo N. (2001) "A Review of Procedures used for the Correction of Seismic data" Paper 39, Civil-Comp Ltd., Striling, Scotland.

Ambraseys, N., Smith, P., Berardi, R., Rinaldis, D., Cotton, F. and Thierry, B.C. (2000) Disseminated of European Strong-Motion Data. CD-ROM Collection. European Council, Environment and Climate Research Programme

Bell and Sejnowski (1995) The maximum entropy method for blind source separation, Computational Neurobiology, La Jolla, CA 92037, USA.

Bermejo, S (2007) Finite sample effects of the fast ICA algorithm, Neurocomputing 71 (2007) 392–399.

Comon, P. (1994) Independent Component Analysis, a New Concept? *SignalProcessing*, 36:287-314

Donoho, D.L. (1992) De-noising by soft thresholding. IEEE Transactions on Information Theory, Stanford University, California. 41(3):613-627

Daubechies, I. (1992) Ten Lectures on Wavelets. Society for Industrial and Applied Mathematics, Philadelphia, PA

Eriksson J., Karvanen J., and Koivunen V. (2000) "Source Distribution Adaptive Maximum Likelihood Estimation of ICA Model" In: Proceeding of the Second International Workshop on Independent Component Analysis. Finland, pp. 227 – 232.

Haykin, S (1999) Neural Networks a Comprehensive Foundation. Prentice Hall, Second edition, pp 544-547

Hyvarinen, A., Karhunen, J. and Oja, E (2000) Independent Component Analysis. John Wiley and sons. Second edition, pp 01-11

Hyvarinen, A., and Oja, E (1997) A Fast Fixed Point Algorithm for ICA, Neural Computing 1997 – 1483-1492.

Igel H., Schreiber U., Flaws A., Schuberth B., Velikoseltsev A., and Cochard A. (2003) "Rotational motions induced by the M8.1 Tokachi-oki earthquake" Vol 32, American Geophysical Union, Sep 25.

Karhunen, J. and Koivunen, V. (2002) Blind separation method based on Pearson system and its extensions. Signal Processing Laboratory, Helsinki University of Tech, Finland. 82(2002) 663-673

Karian, Z.A., Dudewicz, E.J., and McDonald, P. (1996) The Extended Generalized Lambda Distribution System for Fitting Distributions to Data: History, Completion of Theory, Tables, Applications, the`Final Word' on Moment Fits, Communications in Statistics: Simulation and Computation, Vol. 25, No. 3, pp. 611-642

Mutihac, R. and Marc Hulle, V.M. (2003) A Comparative Survey on Adaptive Neural Network Algorithms for Independent Component Analysis. Katholieke Universiteit Leuven, Labo voor Neuro- en Psychofysiologie, Campus Ghastuisberg, B-3000 Leuven, Belgium.

Nigbor R.L. (1994) "Six-Degree-of-Freedom Ground Motion Measurement" Vol 84, No 5, Seismological Society of America, October 1994, pp. 1665-1669.

Tukey, J W. (1960) - A practical relationship between the common transformationof percentages of counts and of amounts. Technical Report 36, Statistical techniques research group. (Princeton University).