Independent Component Analysis Using Maximization of L-Kurtosis

Budhaditya Hazra

Department of Civil and Environmental Engineering, University of Waterloo, 200 University Ave. W., Waterloo, ON, N2L 3G1, Canada

Abstract This paper presents a new approach towards independent component analysis (ICA) for small samples of data, utilizing the linear combination of expectations of order statistics, also termed as L-moments. The main advantage of using L-moments is the relatively low bias in their estimation for small samples compared to the conventional moments. In the present work, arguments leading to kurtosis maximization ICA are first explored and a criterion based on the maximization of L-kurtosis is developed. The optimality criterion based on the extraction of a single source is then assessed. The independent components of the mixture are extracted sequentially using a deflationary approach. The quality of separation of independent components from a mixture is re-interpreted in terms of the distribution parameters of the recovered sources. The robustness of the proposed algorithm is demonstrated through simulation examples of separation of 2-source mixtures, a large-scale problem and a case study from health monitoring of civil structures.

Keywords Independent Component Analysis (ICA), small-samples, L-Kurtosis, Ambient system Identification

1. Introduction

Suppose a system is receiving random input in the form of a set of random variables \( s_1, s_2, \ldots, s_p \). The system output is postulated to be some linear combination of these inputs. Consider that the output is a vector, \( X_1, X_2, \ldots, X_n \), with \( n \) distinct components. The input and output relation can be expressed in a matrix form as

\[
x = Ax
\]

where \( A \) is referred to as a mixing matrix.

A key challenge in this problem is that both the mixing matrix (\( A \)) and input sources (\( s \)) are unknown, such that the problem cannot be solved using the principles of linear algebra alone. It is necessary to impose certain statistical conditions on the nature of the sources (\( s \)). The sources need to be statistically independent of each other and non-Gaussian in nature (at the most one Gaussian source) to enable their estimation using Eq. 1. Finding the sources by the solution of Eq. 1 is often referred to as source separation. The method of separating non-Gaussian and statistically independent components or sources from a mixture of observations is referred to as Independent Component Analysis (ICA). This method is widely used in the signal processing literature[1, 6, 13 and 14].

1.1. Background

The basic idea is to derive a demixing matrix, \( W \), to recover the input independent components, \( s \), as

\[
s = Wx
\]

where the matrix \( W \) is the inverse of \( A \). This inversion is based on maximizing a measure of non-Gaussianity, such as kurtosis one of the foremost measures of non-Gaussianity[6, 15]. Other approaches were later developed that used the concept of maximization of mutual information[14]. The methods falling in this category are FastICA[13,14], Kernel-ICA[3]; RADICAL[19]. The existing algorithms implicitly assume that sufficient data are available to allow highly accurate estimation of a sample statistic, as for example the kurtosis. So a key question is what is the robustness of existing algorithms in extracting the input source vectors?

It is known that a small-sample estimate of kurtosis is highly sensitive to sampling error, i.e., the sample estimate tends to be biased and with a high standard error[12]. This problem can be circumvented if a sample measure of kurtosis is that has relatively less bias and standard error for small samples is used. The use of L-kurtosis for smaller samples of data provides a possible solution to the above problem[12].

1.2. Basic philosophy of ICA

Recovery of sources according to Eq. (2) is based on the concept of maximization of non-Gaussianity. The underlying philosophy of maximization of non-Gaussianity is Central Limit Theorem (CLT). CLT is a classical result from the theory of probability, which states that the distribution of a sum of random variables usually has a distribution that
is closer to Gaussian than any two of the original random variables. A direct consequence of CLT and some of its related extensions[9, 16 and 21] like Carmer’s lemma and Darmois theorem is that, the weighted sum of the sources in Eq. (1) is more Gaussian than each of the sources $S_i$ and becomes least Gaussian when $i = 1$. Since no information about the mixing matrix is known A apriori, the estimation of $W$ is achieved by assuming an initial value and looking for one of the column vectors of $W$ that maximizes the kurtosis of the product vector $Wx$. This facilitates the estimation of one of the components sometimes referred to as one unit separation[13,14].

One of the key steps of maximizing non-Gaussianity is whitening[13,14]. It is essentially a linear transformation that uncorrelates a mixture of random variables, often referred to as principal component analysis[14]. Thus, a combination of whitening and maximization of non-Gaussianity is essential to separate independent components from a mixture of observations.

1.3. Research Objective

ICA has recently been extended to modal identification of civil and mechanical systems[23 and 24]. It basically involves extraction of the information of the dynamic parameters of vibrating systems like buildings, bridges, mostly in absence of the knowledge of inputs. Under certain conditions related to the input spectral characteristics where only output data is available at the disposal, the normal modes of a linear dynamical system can be thought of as being the virtual sources or virtual inputs. Hence, separating these virtual inputs from a mixture of sources then becomes a blind identification problem. In this paper, the problem of modal identification of realistic structural systems under commonly encountered broad-band excitations is addressed though an example using the proposed new ICA method. Although the classical ICA algorithms are known to be robust under most of the practical circumstances in signal processing applications, their performance with smaller sample sizes still remains to be comprehensively studied. To cater to such applications it is imperative to develop ICA algorithms that work fairly robustly for small sample sizes. Moreover, under such circumstances it becomes important to understand to what extent the distribution parameters of the sources can be recovered, a study which hasn’t been conclusively carried out to the knowledge of the authors.

To address the aforementioned problems, an ICA algorithm based on the maximization of L-kurtosis criterion is proposed in this paper. L-kurtosis[12], unlike the conventional kurtosis is estimated by linear combination of order statistics[10], hence it is essentially a linear measure. The use of order statistics in ICA is not new[5, 18, and 22]. In[5], a novel method to perform ICA through order statistics is presented. A Gaussianity measure based on the cumulative density function is introduced. In[18], Pham used order statistics to estimate a discretized form of mutual information in order to perform the separation. In[22], Karvanen used L-Kurtosis for modeling sources using generalized lambda distribution, which was subsequently utilized to form adaptive score models for maximum likelihood type approach to ICA. The main advantages of L-moments are that, being a linear combination of data, they are less influenced by outliers, and the bias of their small sample estimates remains fairly small. A fixed point algorithm following the lines of FAST-ICA is proposed in which the gradient is no longer a non-linear function of $W$, but a constant term that depends on the L-Kurtosis estimated from the whitened data under certain restricted assumptions. This algorithm will be referred to as $L$-kurtosis-ICA henceforth.

1.4. Organization

In this paper, proceeding from the introductory and background discussions, the Kurtosis maximization is discussed in the next section. This is followed by the detailed exposition of the proposed L-kurtosis-ICA algorithm and subsequent numerical studies wherein various case studies are reported in detail. Finally, a practical example of system identification is taken up to show the efficacy of this method. The paper is finally concluded by summarizing the important outcomes and the key points of the present work.

2. Review of Kurtosis Maximization

The first step common to most of the ICA algorithms is called whitening[2,14] which essentially involves rotating the mixture of observations given by Eq. 1 to plane such that the rotated vectors are uncorrelated to each other. It is an orthogonal transformation in which the covariance matrix of $x$ for zero lags, $R_x(0) = (1/N)(\sum_{k=1}^N x(k)x^T(k))$, is diagonalized using eigen value decomposition of the form: $R_x(0) = \Lambda_x V_x V_x^T$, where $V_x$ and $\Lambda_x$ are the eigenvectors and the eigenvalues of the co-variance matrix of $x$ respectively. Then, the standard whitening is realized by a linear relation expressed as under:

$$z = \Lambda_x^{-1/2} V_x^T x$$

(3)

Since nothing is known about the un-mixing maxing $w$ apriori, the second step of the kurtosis maximization algorithm to find suitable weight vector $w$, which proceeds by forming a linear combination of the whitened vector $z$ in the form of $y = w^T z$. The Kurtosis of $y$ can be written as:

$$K = E[(y^4)/E[y^2]^2] - 3$$

(4)

As a consequence of whitening, the weight vector $w$ has unit length then $E[(w^T z)^2] = 1$, so that Kurtosis can be re-written as:

$$K = E[(w^T z)^4] - 3$$

(5)

The gradient of Kurtosis for $y = w^T z$ can be shown to be:

$$\frac{\partial K(w^T z)}{\partial w} = c E[z(w^T z)^3]$$

(6)

where $c$ is a constant of proportionality, which is set to unity for convenience. This gradient changes the length as well as the angle of $w$. If we restrict the length of $w$ to unity then the rule of updating $w$ is given by[15]:

$$w_{new} = w_{old} + \eta E[z(w^T z)^3]$$

(7)

This algorithm will be referred to as kurtosis-ICA for
3. Proposed Method: L-Kurtosis-ICA

The main problem considered in this section is development of an expression L-moments[13] are linear combinations of expectations of order statistics[10], which provides useful alternatives to moments for estimating the location, scale and shape of probability distributions.

3.1. L-Kurtosis

The $r^{th}$ L-moment of a random variable $X$ with cumulative distribution function $F$ and quantile function $Q$ is

$$\lambda_r = E\{X p_{r-1} F(\tilde{X})\} = \int_0^1 p_{r-1}(u) Q(u) \, du$$

(8)

where, $p_{r-1}(\cdot)$ is the $r^{th}$ shifted Legendre polynomial,

$$p_{r-1}(u) = \sum_{k=0}^r (-1)^{r-k} \binom{r}{k} \binom{r}{r+k} u^k$$

In particular, $\lambda_1$, the mean, a location measure, and $\lambda_2$, is a scale measure, equal to one half of Gini’s mean difference. The dimensionless L-moment ratios $\tau_1 = \lambda_1 / \lambda_2$ and $\tau_4 = \lambda_4 / \lambda_2$ are measures of skewness and kurtosis, respectively. L-moments can be used as summary statistics for data samples, and to identify probability distributions that fits with the data[13]. L-moments are widely used in the environmental sciences to summarize data and fit frequency distributions: recent examples include[12, 19]. In [15], Jones and Balakrishnan pointed out some relationships between integrals occurring in the definition of moments and L-moments.

L-moments are related to expected values of order statistics. The order statistic $X_{j:n}$, a random variable distributed as the $j^{th}$ smallest element of a random sample of size n drawn from the distribution of X, has expected value:

$$E(X_{j:n}) = \sum_{j=0}^n \binom{n}{j} u^{-1} (1 - u)^{n-j} Q(u) \, du$$

(9)

The various L-moments, expressed in terms of expectations[12] are as under:

$$\lambda_1 = E(X_{1:1}), \ \lambda_2 = \frac{1}{2} E(X_{2:2} - X_{1:2})$$

and in general

$$\lambda_r = \frac{1}{r} \sum_{j=0}^{r-1} (-1)^{r-j} \binom{r-1}{j} E(X_{r-j:r})$$

(10)

3.2. Proposed Measure of Non-Gaussianity

A quantile based measure of non-gaussianity using L-kurtosis is proposed next. L-kurtosis, exploiting order statistics of sources can be used as a non-parametric estimator of the quantile function. The proposed measure of distance uses the quantile function of the estimated sources and compares it to the quantile function of the Gaussian distribution $N(0,1)$. Although the use of order statistics is not new in ICA[5, 17 and 20], the manner in which it is exploited in the present work is quite different from the previous works. In the present work, the proposed measure of non-gaussianity is used to extract one source and then the deflation approach is employed used to extract the sources sequentially.

To measure the distance of a zero-mean unit variance random variable $x$ to the Gaussian distribution, using the Eq. 8, the following criterion is proposed:

$$D(x) = \int_0^1 p_{r-1}^*(u) [Q_x(u) - Q_N(u)] \, du$$

(11)

where $Q_N(u)$ represents the quantile function of Gaussian distribution with 0 mean and unit variance $N(0,1)$.

$$D(x) = 0 \ if \ and \ only \ if \ x \ is \ Gaussian.$$ Eq. 11 can be estimated in a similar manner as Eq. 10:

$$D(x) = \sum_{j=0}^3 p_{r-1}^*(u) \left( E\{Q_x(u) F(x)^j\} - E\{Q_N(u) F(N)^j\} \right)$$

(12)

3.3. Algorithm for L-kurtosis-ICA

To develop an update rule similar to Eq. 7, an expression for the derivative of the quantile is developed next. Recalling that the proposed method aims at extracting a single source and then applying a deflation based approach to recover all the sources, the estimated source $\hat{s}$ can be represented as a linear combination of sources in the most general form as:

$$\hat{s} = \sum_{i=1}^N w_i s_i$$

(13)

In the neighbourhood of a true source only one of the coefficients $w_i$ will be equal to 1 and the rest of them equal to zero.

The derivative of the distance function for $\hat{s}$ (i.e. $D(\hat{s})$) exists, provided $E[\tilde{s}_i] < \infty$ and atleast one source $s_i$ is independent of others for some value of $i$ (say $i = k$). $D(\hat{s})$. It can be written as:

$$\frac{\partial D(\hat{s})}{\partial w_i} = \sum_{j=0}^3 p_{r-1}^* E[s_i] \hat{s} = Q_x(u) F(s_i)^j + \sum_{j=0}^3 E\left\{Q_x(u) F(s_i)^j \right\} \frac{\partial F(s_i)}{\partial w_i}$$

(14)

Consider a value of $w_i$ in the set $\{w_i\}_{i \in \{1:N\}}$ that does not make the derivative in Eq. 19 zero, but is in the neighborhood of one of the values that makes the derivative zero. The estimate of source $\hat{s}$ that is obtained as an outcome of gradient like algorithm (Eq. 7) doesn’t correspond to a single source, instead, a linear combination of the sources. According to the central limit theorem, the distribution of this estimate is closer to the Gaussian distribution than one of the original sources. Thus the magnitude of the distance function $D(\hat{s})$ decreases as the mixture gets closer to the Gaussian distribution meaning thereby that for the set of values $\{w_i\}_{i \in \{1:N\}}$, the derivative in Eq. 14 is larger than the set that corresponds to a zero value of the criterion gradient. It only corresponds to a zero value when $\hat{s}$ equals one of the sources.

Since the expression of the gradient in Eq. 14 is complicated, it is difficult to prove the existence of local minima for the sets of parameters $\{w_i\}_{i \in \{1:N\}}$. However, in the neighborhood of a single source, considerable simplification can be achieved. Around an expected solution, say $\hat{s}$ equals to $s_k$, the vector of $w_i$ becomes $w = [0 \ldots 0, 1, 0 \ldots 0]$, where non-null values is at the index $k$. The first term in the Eq. 14 becomes:

$$E[s_i] \hat{s}_k = Q_x(u) F(s_k)^j = E\{Q_x(u) F(s_k)^j\}$$

(15)

The second term in Eq. 14 becomes equal to zero. The derivative of the distance function thus becomes:
\[ \frac{\partial D(\hat{f})}{\partial w_i} = \sum_{j=0}^{3} p_{3j}^* E(Q_{s_k}(u)F(z)^j) \]  

(16)

As discussed earlier, the proof of the existence of local minima for the sets of parameters \{\(w_i\)\}_{\{1,2\}} is difficult. Hence, we conjecture that if Eq. 16 is used in the gradient update using \(z\) instead of \(\hat{s}_k\), we can still converge to the local minima under the constraint of whitening. Thus, the gradient update rule becomes:

\[ w_{\text{new}} = w_{\text{old}} + \eta \sum_{j=0}^{3} p_{3j}^* E(Q_{s}(u)F(z)^j) \]  

(17)

4. Numerical Studies

4.1. Background

To validate the robustness and the efficacy of the proposed procedure, simulation experiments are carried out using distributions of gamma family and the results are compared with kurtosis maximization algorithm. Since, the primary interest in this study is to observe the small sample size effects, mixtures are generated considering only 100 samples of sources. 2 case studies are considered, the first one involving 2 sources and the second one a large scale example with many sources. For the 2 cases the results are also compared with the and also the standard ICA algorithm (FAST-ICA). For all the cases of 200 trials of Monte-carlo (MC) simulation are carried out. The distributions of the sources used in the present study are shown in Table 1.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Excess Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rayleigh</td>
<td>3.24</td>
</tr>
<tr>
<td>Exponential</td>
<td>9.00</td>
</tr>
<tr>
<td>Gamma (a=2, b=1)</td>
<td>5.00</td>
</tr>
<tr>
<td>Gamma (a=3, b=1)</td>
<td>6.00</td>
</tr>
<tr>
<td>Uniform</td>
<td>1.80</td>
</tr>
</tbody>
</table>

The performance of any ICA algorithm is frequently expressed in terms of amari-index (AI)[1,2], which essentially compares the recovered unmixing matrix to the mixing matrix. AI is defined as:

\[ AI = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} \left( \frac{g_{ij}}{\max g_{ij}} - 1 \right) \]

\[ + \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} \left( \frac{g_{ij}}{\max g_{ij}} - 1 \right) \]  

(18)

where \(g_{ij}\) is the \((i,j)\) element of the matrix \(G = WA\). The term \(\max g_{ij}\) represents the maximum value among the elements in the \(i\)th row vector of \(G\). Similarly, the term represents \(\max g_{ij}\) does the maximum value among the elements in the \(i\)th column vector of \(G\). When the perfect separation is achieved, the performance index is zero. In practice, a performance index around 0.01 indicates quite a good performance.

4.2. Orthonormal Mixing with Two Sources

In this section, performance of the proposed algorithm for 2-sources mixture is considered. Both the sources are generated from the same distribution. For possible engineering applications of statistics, the focus here is more on leptokurtotic sources belonging to distributions like exponential, gamma(2,1) and gamma(3,1). Exponential sources have an excess kurtosis of 9. For gamma (2,1) and gamma (3,1) they are 6 and 5 respectively.

The sources are mixed according to Eq. 1. The mixing matrix is a fixed orthonormal matrix. Since the probability distribution a random vector is sensitive to scaling the sources are estimated by first recovering the mixing matrix \(A\) and then applying Eq. 2 on the raw un-centered data. Fig. 1 shows the average parameter values (i.e., average of 200 MC trials) and standard error ratio in the parameters of a pair of identified exponential (\(\lambda=1\)) sources. It is quite clear from Fig. 1 that the bias of the estimated parameter of exponential distribution is of the order of 5% for 100 samples of data which increases to 15% for 40 samples if the kurtosis maximization algorithm is used. The L-kurtosis algorithm, however, shows much improvement in the identified parameter values with a bias of 5% for 40 samples of data. A similar picture is also reflected in the standard error values expressed as ratio of the original parameter value which is 1 in this case (i.e., \(\lambda=1\)).

A similar study is carried out for a pair of sources belonging to two parameter family distribution, namely gamma, with shape parameter \(a\), and shape parameter \(b\) commonly referred to as \(\text{gamma}(a, b)\). It is clear from Fig. 2, that the bias in the identified value of \(a\) is of the order of 7.5 % for 100 samples using L-kurtosis ICA and almost 15% if kurtosis-ICA is used. With reduction in sample size to 40, the numbers go up to 11% and 28% for L-kurtosis ICA and kurtosis-ICA respectively. Similar pattern of the bias values is also observed for the identified parameter \(b\). An examination of the standard error values reveals a similar picture. For the identified values of parameter \(a\), the standard error ratio is of the order of 0.2 for using L-kurtosis algorithm for 100 samples. The ratio goes up to 0.27 when the sample size is reduced to 40. The numbers corresponding to kurtosis-ICA algorithm are 0.3 and 0.44 respectively. For parameter \(b\) the standard error ratio shows a significantly less variation.

Thus, it can be safely concluded that the l-kurtosis algorithm is more robust and accurate than the kurtosis-ICA algorithm for sample sizes in the range of 40-100.

To develop insight into the relationship of the amari-index (AI) with the identified values of distribution parameters for small sample-sizes, mean values of AI are plotted for 200 trials of MC simulation with respect to sample sizes for some commonly used distributions (Table 1). It can be seen from the fig. 3 that AI for the kurtosis-ICA method is close to 0.2 for uniform distributed sources and increases to 0.5 for sample size 40. For exponential, the numbers are still higher. The performance worsens for leptokurtotic sources that approach the excess-kurtosis of Gaussian distribution which is 3. The L-kurtosis algorithm performs much better in this regard evidenced from AI values of 0.1 for exponential sources approaching to a value of 0.2 for 40 samples of data.

For gamma (2, 1) it is 0.2 for 100 samples of data and 0.26 for 40 samples. The performance deteriorates for leptoc-
kurtotic sources as they approach an excess kurtosis of 3. It is interesting to note that there is no improvement in the performance of l-kurtosis ICA for a pair of uniformly distributed sources when compared to kurtosis-ICA.

Figure 1. Bias and standard error ratio for a pair of exponential (λ=1) separated sources

Figure 2. Bias and standard error ratio for a pair of gamma (2,1) separated sources
An interesting observation in this study is the indirect relationship of AI and the identified values of distribution parameters from the viewpoint of L-kurtosis ICA algorithm. AI, typically in the order of 0.01 is a standard result for large sample sizes (typically 1000) in blind signal separation. But for small samples it can be seen that AI values of 0.1 and 0.2 in case of exponential distribution corresponds to a bias of 2% and 5.2% in the values of the identified exponential parameters. Similarly AI of 0.2 and 0.26 in case of gamma (2,1) sources corresponds to a bias of 7.5% and 12.5% in the values of the identified values of parameter \(a\). Thus for all practical purposes in the separation of exponential and gamma distributed sources, AI values in the range of 0.1 to 0.2 corresponds to reasonably good estimates of the distribution parameters identified using l-kurtosis ICA algorithm.

4.3. A 5-source Mixture

To assess the efficacy of the algorithm in large scale mixing, a 5-source mixing problem is considered. The sources are sampled from distributions shown in Table 1. They are mixed using a 5x5 randomly generated orthonormal mixing matrix. A set of 200 MC simulations is carried out and mean values of AI for various sample sizes is shown in Fig. 4. It can be observed from the results in Fig. 4 that for the 5-source mixture L-kurtosis ICA performs much better than kurtosis-ICA for small samples as evidenced from the AI values. Whereas for L-kurtosis ICA, AI ranges between 0.3 for 100 samples data to 0.42 for 40 samples, the same numbers kurtosis-ICA are 0.48 for 100 samples and 0.54 for 40 samples. It can also be observed that L-kurtosis ICA performs marginally better than Fastica when compared from the light of AI values. Thus, it can be concluded that when the number of samples are below 100, L-kurtosis performs better than 2 very well known algorithms namely kurtosis ICA and Fastica.

5. Case study: Ambient System Identification

5.1. Simulation Study

In the era of high performance flexible structures estimation of dynamic properties of flexible structures, i.e., natural frequencies and damping, is of paramount importance. Its significance stems from the fact that such structures are liable to undergo changes over time, sustain damage. Extracting system information often needs to be accomplished without the knowledge of inputs, and is referred to as ambient system identification or blind system identification. To understand the concept of ambient system identification, it is instructive to consider the equations of motion for a multi degree of freedom structural system under the action of an excitation force vector \(F(t)\) can be expressed as:

\[
M\ddot{x}(t) + C\dot{x}(t) + Kx(t) = F(t)
\]  

(19)

where, \(x(t)\) is a vector of displacement coordinates at the degree of freedom. Under special cases of the excitation vector \(F\), for example when \(F(t) = 0\), which corresponds to the case of free vibration, the solution to (19) can be written in terms of an expansion of vibration modes as under (in matrix form):
where $x \in \mathbb{R}^{n \times N}$ is the trajectory matrix composed of the sampled components of $x$, and $Q \in \mathbb{R}^{n \times N}$ is a matrix of the corresponding modal coordinates and $\Psi$ the modal transformation matrix. Note the similarity between (20) and (1). Under special circumstances [23], the normal modes or the modal coordinates can be regarded as the most independent sources (termed as virtual sources), thus rendering the presence or the absence of the external force inconsequential. Thus, the modal coordinates $Q$ are a special case of general sources $s$ with time structure. Furthermore, distinct modal coordinates automatically meet the requirement of independence as of sources in ICA and forms the basis of the modal identification procedure [23, 24, 25 and 26].

For the more general case of excitations $F(t)$ (e.g. uncorrelated white noise) it is possible to write the correlation of responses in (19) in the form of (20)[24]. The cross- correlation between the locations denoted by $i$ and $j$ due to an input at $k$ is given by:

$$R_{ijk}(T) = E[x_{ij}(t + T)x_k(t)] \quad (21)$$

Assuming that the disturbance is a white noise process, $E[f_s(\tau_1)f_s(\tau_2)] = \alpha_s(\tau_1 - \tau_2)$ (22) where $\alpha_s$ is a constant and $\delta$ is the dirac-delta function. Under these conditions,

$$R_{ijk}(T) = \sum_{r=1}^{3} A_r e^{-\zeta_r \omega_{dr} T} [\alpha_r \cos \omega_{dr} T + \beta_r \sin \omega_{dr} T]$$

where $A_r$ contains information about the modes and the constants $\alpha_r$ and $\beta_r$ depict the decay due to damping for the details of which the readers are referred to [24].

The quantity $\lambda = t - \tau$, and $\omega_d$ represents the damped natural frequency. For the $r^{th}$ mode, $\omega_d = \omega_u \sqrt{1 - \zeta_r^2}$, where $\omega_u, \zeta_r$ are the undamped natural frequency and the critical damping in the $r^{th}$ mode. Making the substitution, $s_r = e^{-\zeta_r \omega_{dr} T} [\alpha_r \cos \omega_{dr} T + \beta_r \sin \omega_{dr} T]$ in matrix form, eq. 23 becomes,

$$x(t) = R(t) = As(t) \quad (24)$$

In the above form, it is easy to recognize the similarity between (20) and (24), provided $R$ is substituted for $x$. For the purposes of this study, the vector $R$ is used instead of the structural responses, $x$. Hence, under the cases of free vibration and broadband white excitations, the problem of modal identification can be cast into the framework of ICA, where-in the modes represent the independent sources and the modal coordinates is the mixing matrix [24, 25 and 26].

In order to conduct parametric and comparison studies, numerical simulations are carried out on a simple three-degree of freedom (3DOF) mass-spring-dashpot system. For convenience in studying the effect of damping, the damping is assumed to be mass proportional. The state equations for this system subjected to an external excitation vector $w$ can be written as:

$$\dot{x} = Ax + Ew \quad y = Cx \quad (25)$$

Here, the vector $x$ is a vector of states, and the vector $y$ represents the output vector, which is governed by the $C$ matrix. The system matrix, $A$, and the excitation matrix $E$, $M$ and $C$ are shown in the Figure 5.

$$A = \begin{bmatrix} 0 & 1 \ -M^{-1}K & -M^{-1}C \end{bmatrix}$$

$$E = \begin{bmatrix} 0 & 0 & 0 \ -1 & -\frac{1}{m} & -\frac{1}{m} \end{bmatrix}$$

$$M = \begin{bmatrix} m & 0 & 0 \\ 0 & m & 0 \\ 0 & 0 & m \end{bmatrix} ; C = \begin{bmatrix} c \ 0 \ 0 \ 0 \ c \ 0 \ 0 \ c \ 0 \ 0 \ c \ 0 \ 0 \ c \ 0 \ 0 \ c \end{bmatrix}$$

Figure 5. System Matrices

where, $m$ represents the mass at each floor level which is taken to be 15 kgs, $c$ the damping coefficient, and $k_1, k_2, k_3$ correspond to the linear stiffness at each degree of freedom. Constant stiffness of $2k/N/m$ is used for the springs and $c = 0.01\eta m$ is used. The variable $\eta$ is used to vary the level of damping to study the performance of the BSS methods. The correlation between the vibration modes and BSS modes is performed using the modal assurance criterion (MAC)[24,25]. MAC values range between 0 and 1, a value of 1 meaning a perfect correlation. MAC is defined as:

$$MAC_i = (\psi_i^T \Psi_i )^2 / (\psi_i^T \psi_i ) (\psi_i^T \Psi_i ) \quad (26)$$

Here, $\psi_i$ and $\Psi_i$ represent the $i^{th}$ theoretical and the estimated mode shape vectors respectively. For the simulation study, a synthetic Gaussian stationary white noise with zero mean and unit standard deviation is considered. A sampling frequency of 100 Hz is considered for the simulation study. The effect of increasing the damping parameter $\eta$ is studied first, and the results displayed in Table 2. The results of kurtosis-ICA, L-kurtosis ICA and Stochastic Sub-space Iteration (SSI)[24, 25, and 26] method are presented. SSI is included in this study as a point of reference for the discussed methods, as SSI has been used extensively in the structural system identification literature[24,26]. The effect of damping is achieved by gradually increasing the value of $\eta$ from 10 to 30 which corresponds to a range of 0.8 % -2.5% critical in the first mode. From the results in Table 2, it is clear that Kurtosis-ICA can identify structures only for small levels of damping, that is for $\eta = 18$, which correspond to 1.2 % critical damping, beyond which the accuracy of identification reduces. L-Kurtosis-ICA shows much improved results up to 2.5% damping. Comparative results between SSI and L-kurtosis-ICA show that for larger sample sizes the accuracy of identification are almost same up to 2.5% damping. But L-Kurtosis ICA scores over SSI when the sample sizes are reduced to 500. MAC values using L-kurtosis ICA are still more than 0.9 where as for SSI the MAC fall down below 0.8. Thus the improved identification performance of L-Kurtosis ICA over SSI and Kurtosis-ICA is one of the key results.

5.2. Experimental Study

In order to demonstrate the practical application of this method, a simple experiment is performed to identify the translational frequencies and modes of a two-story model.
The model shown in Fig. 6 consists of two steel plates of dimensions $12 \times 12 \times 0.75$ in to serve as floor masses, four $0.5$ in aluminum equal angles to serve as columns, of total height $56$ in and $0.0625$ in thick, continuous between both the floors. The top story is braced in one direction using aluminum bars of dimensions $0.125 \times 0.375 \times 0.5$ in. The model is instrumented at both the floor levels, in both the directions, using four accelerometers. The structure is subjected to impact excitation using a rubber mallet, and the acceleration data is collected using DSPACE1104 DAQ and control board. The sampling frequency is set at 100 Hz. The recorded acceleration responses are processed using both L-Kurtosis ICA method and SSI method without performing any post-processing on the data (such as de-noising, etc.). The number of samples of data considered is 2000. The results of the identification are presented in Table 3.

The power spectral density plot of the recorded response of the second floor is shown in Fig.7. The MAC numbers comparing SSI and L-Kurtosis ICA methods show that the quality of identification is similar in both these methods and the identified frequencies correspond well with the peaks in the PSD estimate shown in Fig. 7. The natural frequencies are estimated by fitting a single degree of freedom system to the estimated sources using curve-fitting techniques. This shows that the L-Kurtosis ICA method is well equipped to handle experimental data as well.

Table 2. Comparison of Kurtosis-ICA, L-Kurtosis-ICA and SSI methods for various $\eta$

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>$\eta$</th>
<th>Kurtosis-ICA</th>
<th>L-Kurtosis ICA</th>
<th>SSI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mac1</td>
<td>Mac2</td>
<td>Mac3</td>
</tr>
<tr>
<td>2000</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>0.77</td>
<td>0.64</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1000</td>
<td>10</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>0.72</td>
<td>0.62</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>500</td>
<td>10</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.97</td>
<td>0.95</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>0.71</td>
<td>0.59</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 3. System identification of the experimental model

<table>
<thead>
<tr>
<th>Mode #</th>
<th>Identified Frequency(Hz)</th>
<th>Identified Frequency(Hz)</th>
<th>MAC (SSI—L Kurtosis ICA)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L-Kurtosis ICA</td>
<td>SSI</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.37</td>
<td>1.41</td>
<td>0.96</td>
</tr>
<tr>
<td>2</td>
<td>1.46</td>
<td>1.41</td>
<td>0.97</td>
</tr>
<tr>
<td>3</td>
<td>4.74</td>
<td>4.75</td>
<td>0.99</td>
</tr>
<tr>
<td>4</td>
<td>32.16</td>
<td>32.17</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Figure 6. Experimental set-up

Figure 7. Power spectrum of the top floor response
6. Conclusions

In this paper a new ICA algorithm based on order statistics is introduced. A gradient algorithm based on the maximization of L-Kurtosis is developed. Simulation examples show that the algorithm works quite robustly for small samples of data and gives consistently better results than kurtosis-ICA. For a mixture of 2 exponential sources (λ=1) the bias in estimating λ is only 5% by the use of L-kurtosis-ICA which is as high as 15% if kurtosis-ICA is used for sample size 40. For a pair of Gamma (a=2, b=1), the bias in the identified value of a is 11% for L-kurtosis ICA and 28% for and kurtosis-ICA respectively. For the identified values of parameter a, the standard error ratio is 0.3 when the sample size is reduced to 40 for L-Kurtosis-ICA and 0.42 for kurtosis-ICA. For parameter b the standard error ratio shows a significantly less variation. For the 5-sources mixture the mean AI values for L-kurtosis ICA ranges between 0.3 for 100 samples data to 0.42 for 40 samples, the same numbers kurtosis-ICA are 0.48 for 100 samples data and 0.54 for 40 samples.

In the practical example of an ambient system identification problem, L-kurtosis shows promising results. Clearly, Kurtosis-ICA, which is the most popular method to perform BSS fails to perform adequately in small sample sizes. In this regard, L-Kurtosis ICA performs much better and even outperforms SSI for smaller sample sizes for moderate levels of damping. The ability to identify modal information even at smaller sample sizes demonstrates that the L-Kurtosis ICA method holds significant promise in the area of structural system identification. Preliminary experimental results show that the algorithm is capable of performing satisfactorily under practical situations.

ACKNOWLEDGEMENTS

The author would like to thank the Dr. Mahesh D. Pandey, the chair of the NSERC-UNENE-Industrial Research program (University of Waterloo, CANADA) for the financial support to conduct the study.

REFERENCES

