A Review of Independent Component Analysis Techniques and their Applications

D. P. Acharya and G. Panda

Department of Electronics and Communication Engineering, National Institute of Technology, Rourkela-769 008, India.

Abstract

Independent Component Analysis, a computationally efficient blind statistical signal processing technique, has been an area of interest for researchers for many practical applications in various fields of science and engineering. The present paper attempts to treat the fundamental concepts involved in the independent component analysis (ICA) technique and reviews different ICA algorithms. A thorough discussion of the algorithms with their merits and weaknesses has been carried out. Applications of the ICA algorithms in different fields of science and technology have been reviewed. The limitations and ambiguities of the ICA techniques developed so far have also been outlined. Though several articles have reviewed the ICA techniques in literature, they suffer from the limitation of not being comprehensive to a first time reader or not incorporating the latest available algorithm and their applications. In this work, we present different ICA algorithms from their basics to their potential applications to serve as a comprehensive single source for an inquisitive researcher to carry out his work in this field.

Keywords:
Blind source separation, Higher order statistics, Independent component analysis.

1. Introduction to ICA

Recently, there has been an increasing interest in statistical models for learning data representations. A very popular method for this task is independent component analysis (ICA), the concept of which was initially proposed by Comon [1]. The ICA algorithm was initially proposed to solve the blind source separation (BSS) problem i.e. given only mixtures of a set of underlying sources, the task is to separate the mixed signals and retrieve the original sources [2,3]. Neither the mixing process nor the distribution of sources is known in the process. A simple mathematical representation of the ICA model is as follows.

Consider a simple linear model which consists of N sources of T samples i.e. $\mathbf{s}_i = [s_i(1)...s_i(t)...s_i(T)]$. The symbol t here represents time, but it may represent some other parameter like space. M weighted mixtures of the sources are observed as X, where $X_i = [X_i(1)... X_i(t)... X_i(T)]$. This can be represented as -

$$X = A S + n \tag{1}$$

where

$$X=\begin{bmatrix} X_1 \\ X_2 \\ \dots \\ X_M \end{bmatrix} S=\begin{bmatrix} S_1 \\ S_2 \\ \dots \\ S_N \end{bmatrix} and\ n=\begin{bmatrix} n_1 \\ n_2 \\ \dots \\ n_K \end{bmatrix} \tag{2}$$

S and n represent the additive white Gaussian noise (AWGN). It is assumed that there are at least as many observations as sources i.e. M ≥ N. The M × N matrix A is represented as -

$$A=\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \dots & \dots & \dots & \dots \\ a_{M1} & a_{M2} & \dots & a_{MN} \end{bmatrix} \tag{3}$$

A relates X and S. A is called the mixing matrix. The estimation of the matrix S with knowledge of X is the linear source separation problem. This is schematically shown in Figure 1.

The source separation problem cannot be solved if there is no knowledge of either A or S, apart from the observed mixed data X. If the mixing matrix A is known and the additive noise *n* is negligible, then the original sources can be estimated by evaluating the pseudo inverse of the matrix A, which is known as the unmixing matrix B,

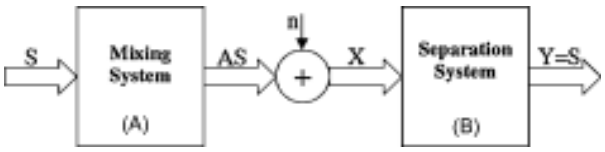

Figure 1: Illustration of mixing and separation system. (A) is the mixing matrix and (B) is the unmixing matrix.

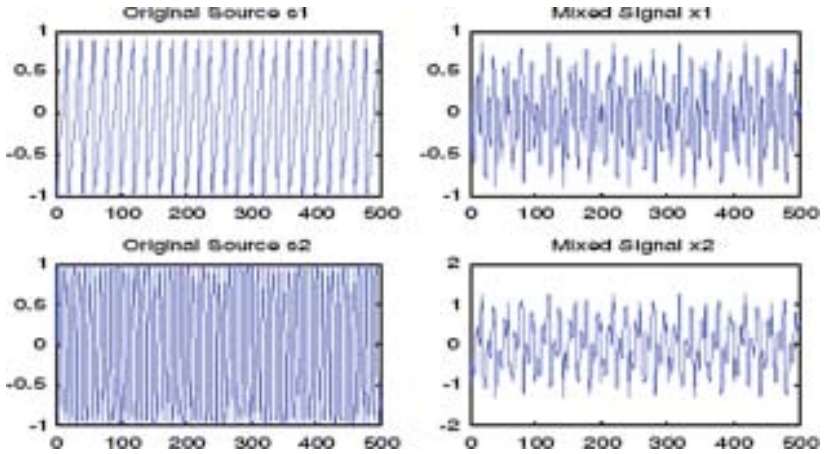

Figure 2: Effect of mixing. The original sources s_1 and s_2 are shown in left plot, and the mixed signals x_1 and x_2 are shown in the right plot.

such that

$$BX = BAS = S \tag{4}$$

For cases where the number of observations M equals the number of sources N (i.e. M = N), the mixing matrix A is a square matrix with full rank and $B = A^{-1}$.

The necessary and sufficient condition for the pseudo-inverse of *A* to exist is that it should be of full rank. When there are more observations than the sources (i.e. M > N), there exist many matrices B which satisfy the condition BA = I. Here the choice B depends on the components of *S* that we are interested in. When the number of observations is less than the number of sources (i.e. M < N), a solution does not exist, unless further assumptions are made. On the other side of the problem, if there is no prior knowledge of the mixing matrix A, then the estimation of both A and S is known as a blind source separation (BSS) problem. A very popular technique for solution of a BSS problem is independent component analysis [4]. Estimation of the underlying independent sources is the primary objective of the BSS problem. The problem defined in (1), under the assumption of negligible Gaussian noise **n,** is solvable with the following restrictions:

- The sources (i.e. the components of S) are statistically independent.
- At most, one of the sources is Gaussian distributed.
- The mixing matrix is of full rank.

From the above discussion, the following remarks can be made on ICA.

Remark1: Independent component analysis (ICA) is a linear transformation S = WX of a multivariate signal X, such that the components of S are as independent as possible in the sense of maximizing some objective function $f(s_1,..., s_N)$, which is a measure of statistical independence.

Remark2: ICA can be defined as a computationally efficient statistical signal processing technique for separating a multivariate signal into its components, assuming that all of these components are statistically independent.

2. Statistical Independence

The above discussions make it clear that statistical independence is the key foundation of independent component analysis (ICA). For the case of two different random variables x and y, x is independent of the value of y, if knowing the value of y does not give any information on the value of x. Statistical independence is defined mathematically in terms of the probability densities as - the random variables x and y are said to be independent, if and only if

$$p_{x,y}(x, y) = p_x(x)p_y(y) \tag{5}$$

where $p_{x,y}(x,y)$ is the joint density of x and y, $p_x(x)$ and $p_y(y)$ are marginal probability densities of x and y respectively. Marginal probability density function of x is defined as

$$p_x(x) = \int p_{x,y}(x, y)dy \tag{6}$$

Generalizing this for a random vector $s = [s_1,..., s_N]^T$ with multivariate density p(s) has statistically independent components, if the density can be factorized as

$$p(s) = \prod_{i=1}^{N} p_i(s_i) \tag{7}$$

In other words, the density of s_1 is unaffected by s_2 when two variables s_1 and s_2 are independent. Statistical independence is a much stronger property than uncorrelatedness, which takes into account second order statistics only. If the variables are independent, they are uncorrelated; but the converse is not true.

3. Contrast Functions for ICA

The data model for independent component analysis is estimated by formulating a function which is an indicator of independence in some way and then minimizing or maximizing it. Such a function is often called a contrast function or cost function or objective function. The optimization of the contrast function enables the estimation of the independent components. The ICA method combines the choice of an objective function and an optimization algorithm. The statistical properties like consistency, asymptotic variance, and robustness of the ICA technique depend on the choice of the objective function and the algorithmic properties like convergence speed, memory requirements, and numerical stability depends on the optimization algorithm. The contrast function in some way or the other is a measure of independence. In this section, different measures of independence, which are frequently used as contrast functions for ICA are discussed.

3.1 Measuring Nongaussianity

3.1.1 Central Limit Theorem

The central limit theorem is the most popular theorem in statistical theory and plays a predominant role in ICA. According to it, let

$$x_k = \sum_{i=1}^{K} Z_i \tag{8}$$

be a partial sum of sequence $\{z_i\}$ of independent and identically distributed random variables z_i. Since the mean and variance of x_k can grow without bound as $k \to \infty$, consider the standardized variables y_k instead of x_k,

$$y_k = \frac{x_k - m_{xk}}{\sigma_{xk}} \tag{9}$$

where m_{xk} and σ_{xk} are mean and variance of x_k. The distribution of y_k converges to a Gaussian distribution with zero mean and unit variance when $k \to \infty$.

This theorem has implicit consequences in ICA and BSS. A typical mixture or component of the data vector **x** is of the form

$$x_i = \sum_{j=1}^{M} a_{ij} S_j \tag{10}$$

where a_{ij}, j = 1,..., M are constant mixing coefficients and s_j, j = 1,..., M are the M unknown source signals. Even for a fairly small number of sources, the distribution of the mixture x_k is usually close to Gaussian. In a very simple way, the central limit theorem can be stated to be the sum of even two independent identically distributed random variables that are more Gaussian than the original random variables. This implies that independent random variables are more nongaussian than their mixtures. Hence, nongaussianity is a measure of independence. This is one of the bases of independent component analysis.

3.1.2 Kurtosis

The Central limit theorem discussed above provides a good intuitive idea that nongaussianity is a measure of independence. The first quantitative measure of nongaussianity is kurtosis, which is the fourth order moment of random data. Given some random data **y,** the kurtosis of **y** denoted by kurt(y) is defined as

$$kurt\ (y) = E\{y^4\} - 3E\{y^2\} \tag{11}$$

where $E\{.\}$ is the statistical expectation operator. For simplicity, if we assume y to be normalized so that the variance is equal to unity i.e. $E\{y^2\} = 1$, then kurt $(y) = E\{y^4\} - 3$. This indicates that kurtosis is simply the normalized version of the fourth moment $E\{y^4\}$.

For a Gaussian **y**, the fourth moment equals to $3(E\{y^2\})^2$. So for Gaussian random variables, the kurtosis value is zero and for nongaussian random variables the kurtosis value is non-zero. It may particularly be noted that when the kurtosis value is positive, the random variables are called *supergaussian* or *leptokurtic* and when the kurtosis value is negative, the random variables are called *subgaussian* or *platykurtic*. Supergaussian random variables have a 'spiky' probability density function, with heavy tails, and subgaussian random variables have a flat probability density function. Nongaussianity is measured by the absolute value of kurtosis. The square of kurtosis can also be used. These measures are zero for a Gaussian variable and greater than zero for most nongaussian random variables. As the value of kurtosis goes away from zero, the distribution moves away from the Gaussian distribution i.e. it becomes more nongaussian. However, kurtosis is very sensitive to outliers in data set and this is a limitation of kurtosis as the contrast function.

3.1.3 Negentropy

A second optimal quantitative measure of nongaussianity is negentropy which is based on the information theoretic differential entropy. The entropy of data is related to the information that is observed. The more random and unpredictable the data is, the larger entropy it will have. The entropy S of a random variable y with a density of $p_{(\eta)}$ is

$$S(y) = -\int P_x(\eta) log p_y(\eta) dp_y(\eta) \tag{12}$$

A fundamental result of information theory is that a Gaussian variable has the largest entropy among all random variables of equal variance. This means that entropy could be used as a measure of nongaussianity. This shows that the Gaussian distribution is the 'most random' or least structured of all distributions. Entropy is small for distribution that is clearly concentrated on certain values i.e. when the variable is clearly clustered or has a pdf that is very 'spiky' which means the distribution is away from Gaussian distribution. A measure of nongaussianity that is zero for a Gaussian variable and always nonnegative is obtained by using a normalized version of differential entropy called negentropy. Thus, negentropy is the maximum for nongaussian random variable. Negentropy of **y** denoted by H(y) is defined as

$$H(y) = S(y_{gauss}) - S(y) \tag{13}$$

where $\mathbf{y}_{\mathbf{gauss}}$ is a Gaussian random variable with the same correlation and covariance as **y**. Since the negentropy is normalized, it is always nonnegative and zero if **y** is Gaussian distributed. Negentropy has the additional interesting property that it is invariant for invertible linear transformations.

3.1.4 Approximations to Negentropy

However, negentropy is practically difficult to compute and requires complex computation. A method of approximating negentropy is using higher order cumulants using polynomial density expansions. Using Gram-Charlier expansion in the pdf of y, the following approximation for negentropy results may be worked out:

$$H(y) = \frac{1}{2}E\{y^3\}^2 + \frac{1}{48}kurt(y)^2 \tag{14}$$

This approximation often leads to the use of kurtosis as a contrast function. If some nonquadratic function G is used, then the approximations to negentropy in terms of expectation of the function G is expressed as -

$$H(y) = K[E\{G(y_i)\} - E\{G(v)\}]^2 \tag{15}$$

where K is a constant and v is a Gaussian variable of zero mean and unit variance. Wise choice of G makes a good contrast function H(y) for optimization in ICA. It may be particularly noted that if G is chosen such that it does not grow too fast, then more robust estimators are obtained. The frequent choices of G that have proved useful are:

$$G_1(y) = \frac{1}{a_1}\log\cosh(a_1 y) \tag{16}$$

$$G_2(y) = -\frac{1}{a_2}exp(-a_2 y^2/2) \tag{17}$$

$$G_3(y) = \frac{1}{4}y^4 \tag{18}$$

where a_1 and a_2 are constants. The choice of G as in (18) makes negentropy approximated to kurtosis-based cost function. Under the approximation,

$$E\{(w_I^T z)(w_I^T)\} = \delta_u \tag{19}$$

where δ is known as Kronecker delta function.

H(y) expression in equation simplifies to

$$H(w) = E\{G(w^T z)\} \tag{20}$$

which is a good contrast function for optimization in ICA problems.

3.2 Mutual Information

Mutual information is a natural measure of dependency between random variables i.e. it is a measure of the information that a member of a set of random variables has on the other random variable in the set.

If y is a n-dimensional random variable and $p_y(\eta)$ its probability density function, then vector y has mutually independent components, if and only if

$$p_y(\eta) = p_{y1}(\eta_1)p_{y2}(\eta_2)...p_{y3}(\eta_n) \tag{21}$$

A natural way of checking whether **y** has ICs is to measure a distance between both sides of the above equation.

$$I(p_y) = \delta(p_y, \prod p_{yi}) \tag{22}$$

The average mutual information of **y** as given by Comon [5] as p(η)

$$I(p_y) = \int p_y(\eta)\log\left(\frac{p_y(\eta)}{\prod p_y(\eta)}\right)d\eta \tag{23}$$

The average mutual information vanishes if and only if the variables are mutually independent and are otherwise strictly positive. In terms of negentropy, mutual information is written as -

$$I(y_1, y_2, ..., y_n) = H(y) - \sum H(y_i) \tag{24}$$

But the contrast functions based on mutual information discussed above require the estimation of the density function and this has severely restricted the use of these contrast functions.

Before these optimization functions are used in the ICA optimization algorithm, the observed data is processed as described in the following section:

4. Preprocessing of Data for ICA

Generally, ICA is performed on multidimensional data. This data may be corrupted by noise, and several original dimensions of data may contain only noise. So if ICA is performed on a high dimensional data, it may lead to poor results due to the fact that such data contain very few latent components. Hence, reduction of the dimensionality of the data is a preprocessing technique that is carried prior to ICA. Thus, finding a principal subspace where the data exist reduces the noise. Besides, when the number of parameters is larger, as compared to the number of data points, the estimation of those parameters becomes very difficult and often leads to over-learning. Over learning in ICA typically produces estimates of the independent components that have a single spike or bump and are practically zero everywhere else [5]. This is because in the space of source signals of unit variance, nongaussianity is more or less maximized by such spike/bump signals.

Apart from reducing the dimension, the observed signals are centered and decorrelated. The observed signal **X** is centered by subtracting its mean:

$$X \leftarrow X - E\{X\} \tag{25}$$

Second-order dependences are removed by decorrelation, which is achieved by the principal component analysis (PCA) [6,7]. The ICA problem is greatly simplified if the observed mixture vectors are first whitened. A zero-mean random vector $\mathbf{z} = (z_i.....z_j)^T$ is said to be white if its elements z are uncorrelated and have unit variances $E\{z_i z_j\} = \delta_{i,j}$

In terms of Covariance matrix, the above equation can be restated as,

$$E\{zz^T\} = I \tag{26}$$

where I is the identity matrix. A synonymous term for white is sphered. If the density of the vector **z** is radially symmetric and suitably scaled, then it is sphered, but the converse is not always true, because whitening is essentially decorrelation followed by scaling, for which the PCA technique can be used.

The problem of whitening: Given a random vector x with n elements, we have to have a linear transformation V into another vector z such that

$$z = V x \tag{27}$$

is white or sphered.

Suppose $E = [e_{1.....} e_n]$ is the matrix whose columns are the unit-norm eigenvectors of the covariance matrix $C_x = E\{xx^T\}$ and $D = diag[d_1.....d_n]$ is the diagonal matrix of the eigenvalues of C_x then $C_x = ED E^T$. This is called the eigenvectors decomposition of the covariance matrix. The linear whitening transform is expressed as

$$V = D^{-½}E^T \tag{28}$$

Hence $V = D^{-½}E^T x$ (29)

An ICA estimation is now performed on the whitened data z, instead of the original data x. For whitened data, it is sufficient to find an orthogonal separation matrix, if the independent components are assumed white.

Dimensionality reduction by PCA is carried on by projecting the *N* dimensional data to a lower dimensional space spanned by m (m < N) dominant eigenvectors (i.e. eigenvectors corresponding to large eigenvectors) of the correlation matrix C_x. The eigenvectors matrix *E* and the diagonal matrix of eigenvectors D are of dimension N × m and m × m respectively. Practically, it is a nontrivial task to identify the lower dimensional subspace properly. For noise free data, a subspace corresponding to the nonzero eigenvalues is required to be found. In most of the scenario, data are corrupted by noise and are not contained exactly within the subspace. In this case, the eigenvectors corresponding to the largest eigenvalues should describe the data well; however, in general, 'weak' independent components may be lost in the dimension reduction process. This involves a hit and trial process.

Dimensionality reduction can also be accomplished by methods other than PCA. These methods include local PCA [8] and random projection. For noise reduction, another popular technique called principal factor analysis [9] is used.

The unmixing matrix B in Figure1 can be regarded as a two step process i.e. whitening and rotation. Hence,

$$B = W^T V \tag{30}$$

The whitening matrix $V = D^{-½} E^T$ is estimated by PCA and rotation matrix W is found by one of the ICA techniques described in the following section.

Once the data are whitened, the matrix W is necessarily orthogonal. This reduces the number of parameters to be estimated and enables the use of efficient optimization techniques. The fact that W is an orthogonal matrix in the ICA problem endows the parameter space with addi-

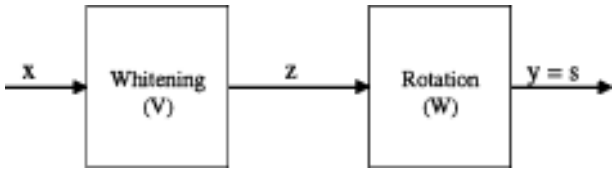

Figure 3: Schematic of separation: whitening and rotation. The unmixing matrix B in Figure 1 can be regarded as a concatenation of the whitening matrix V and the (orthogonal) rotation matrix W

tional structure, which can be exploited by optimization algorithm. This process can be depicted in Figure 3.

5. Algorithms for ICA

Some of the ICA algorithms require a preprocessing of data *X* and some may not. Algorithms that need no preprocessing (centering and whitening) often converge better with whitened data. However, in certain cases, if it is necessary, then sphered data *Z* is used. There is no other mention of sphering done for cases where whitened is not required.

5.1 Non-linear Cross Correlation based Algorithm

The principle of cancellation of nonlinear cross correlation is used to estimate independent components in [10,11]. Nonlinear cross correlations are of the form $E\{g_1(y_i)g_2(y_j)\}$, where g_1 and g_2 are some suitably chosen nonlinearities. If y_i and y_j are independent, then these cross correlations are zero for y_i and y_j, having symmetric densities. The objective function in such cases is formulated implicitly and the exact objective function may not even exist. Jutten and Herault, in [5], used this principle to update the nondiagonal terms of the matrix W according to

$$\Delta W_{ij} \propto g_1(y_i)g_2(y_j) \quad \text{for } i \neq j \tag{31}$$

Here y_i are computed at every iteration as $y = (I + W)^{-1}z$ and the diagonal terms W_{ij} are set to zero. After convergence, y_i give the estimates of the independent components. However, the algorithm converges only under severe restrictions [12].

5.2 Nonlinear Decorrelation Algorithm

To reduce the computational overhead by avoiding matrix inversions in the Jutten-Herault algorithm and to improve stability, some algorithms have been proposed in [13]. Among these, the following algorithm has been proposed:

$$\Delta W \propto (I - g_1(y)g_2(y^T))\, W \tag{32}$$

where $y = Wx$, the nonlinearities $g_1(.)$ and $g_2(.)$ are applied separately on every component of the vector *y*, and the identity matrix can be replaced by any positive definite diagonal matrix. The equivariant adaptive separation via independence (EASI) algorithm has been proposed in [14,15].

According to EASI,

$$\Delta W \propto (I - yy^T - g(y)\, y^T + yg(y^T))\, W \tag{33}$$

The choice of the nonlinearities used in the above rules is generally provided by the maximum likelihood (or infomax) approach.

5.3 Infomax Estimation or Maximum Likelihood Algorithm

Maximum likelihood (ML) estimation is based on the assumption that the unknown parameters to be estimated θ are constants or no prior information is available on them. When the number of samples is large, ML estimator becomes a desirable choice owing to its asymptotic optimality properties. The ML estimation can be simply interpreted as follows: those parameters having the highest probability for the observations act as the estimates. The simplest algorithm for maximizing the likelihood (also log-likelihood) is given by Bell and Sejnowski [16] by using stochastic gradient methods. The algorithm for ML estimation derived by Bell and Sejnowski in [16] is

$$\Delta W \propto [W^T]^{-1} + E\{g(Wx)x^T\} \tag{34}$$

Here the nonlinearity *g* is very often chosen as tanh function because it is the derivative of log density of the logistic distribution. This function works for estimation of most super-gaussian independent components; however, other functions should be used for subgaussian independent components. The convergence of the algorithm described by the above equation is very slow, especially due to the inversion of the matrix *W* that is needed at every step. The convergence of the algorithm can be improved by whitening the data and by using the natural gradient.

The natural (or relative) gradient method simplifies the maximization of the likelihood and makes it better conditioned. The natural gradient principle is based on the geometrical structure of parameter space. This is related to the relative gradient principle, which uses the Lie group structure of the ICA problem. In the case of basic ICA, both these principles amount to multiplying the right side of the above equation by W^TW. This gives

$$\Delta W \propto (I + E\{g(y)y^T\})\, W \tag{35}$$

where y = Wx. After this modification, the algorithm needs no sphering. This algorithm can be interpreted as a special case of nonlinear decorrelation algorithm, which has been described in previous section.

A Newton method for maximizing the likelihood has been introduced in [26]. Though it converges with less iteration, it suffers from the problem that a matrix inversion is needed in each iteration.

Infomax principle [16] is a very closely related to maximum likelihood estimation principle for ICA [17]. This is based on maximizing the output entropy or information flow of a neural network with nonlinear outputs. Hence, it is named as infomax.

5.4 Nonlinear PCA Algorithm

Another approach to ICA that is related to PCA is the so-called nonlinear representation, which is sought for the input data that minimizes a least mean square error criterion. For, linear case principal components are obtained, and, in some cases, the nonlinear PCA approach gives independent components instead. In [10], the following version of a hierarchical PCA learning rule is introduced.

$$\Delta w \propto g(y_i)x - g(y_i)\sum_{j=1}^{i} g(y_i)w_j \tag{36}$$

where g is a suitable nonlinear scalar function. The introduction of nonlinearities means that the learning rule uses a higher order information in the learning. In [18], it is proven that for well chosen non-linearities, the learning rule in the above equation does indeed perform ICA, if the data is whitened. Algorithms for exactly maximizing the nonlinear PCA criteria are introduced in [11].

5.5 One-unit Neural Learning Rules

Simple algorithms from the one-unit contrast functions can be derived using the principle of stochastic gradient descent. Considering the whitened data, Hebbian like learning rule [19, 20] is obtained by taking instantaneous gradient of contrast function with respect to w. The rule is

$$\Delta w \propto [E\{G(w^T x)\} - E\{Gv)\}] \times g(w^T x) \tag{37}$$

Such one-unit algorithms were first introduced in [21] using kurtosis. For estimation of several independent components of system of several units such one-unit algorithms are needed.

5.6 Tensor based ICA Algorithm

Another approach for the estimation of independent components consists of using higher-order cumulant tensors. Tensors are generalizations of matrices, or linear operators. Cumulant tensors are generalizations of the covariance matrix C_x. The covariance matrix is the second order cumulant tensor, and the fourth order tensor is defined by the fourth-order cumulants as $\text{cum}(x_i, x_j, x_k, x_l)$.

Eigenvalue-Decomposition (EVD) is used to whiten the data. Through whitening, the data is transformed so that its second-order correlations are zero. This principle can be generalized so that the off-diagonal elements of the fourth-order cumulant tensor can be minimized. This kind of (approximate) higher-order decorrelation results in a class of methods for ICA estimation.

Joint approximate diagonalization of eigenmatrices (JADE) proposed by Cardoso [6] is based on the principle of computing several cumulant tensors $F(M_i)$, where F represents the cumulant tensor and M_i represents the eigenmatrices. These tensors are diagonalized jointly as well as possible. If a matrix W diagonalizes F(M) for any M, then W F(M) W^T is diagonal since the matrix F is a linear combination of the T terms $w_i w_{i,}$ assuming that the ICA model in equation (1) holds. A measure of the diagonality of $Q = WF(M_i)W^T$ is the sum of squares of the off-diagonal elements $\sum_{kl \neq 1} q^2$. In other words, since the matrix W is orthogonal and it does not change the total sum of squares of a matrix, the minimization of the sum of the squares of the off-diagonal elements is equivalent to the maximization of the squares of the diagonal elements. Thus, the following function can be a good measure of the joint diagonalization process.

$$J_{JADE}(W) = \sum \| diag(WF(M_1)W^T) \|^2 \tag{38}$$

This represents the sum of the squares of all the diagonal elements of all the diagonalized cumulant tensors.

M_i are chosen as the eigenmatrices of the cumulant tensor because the n eigenmatrices span the same subspace as the cumulant tensor, and, hence, they contain all the relevant information on the cumulants. With this choice, the contrast function expressed in the above equation can be restated as -

$$J_{JADE}(W) = \sum_{ijkl \neq iikl} cum\left(y_i, y_j, y_k, y_1\right)^2 \tag{39}$$

where y is the estimate of the independent sources obtained as y = Wx. The above equation means that by minimizing J_{JADE}, the sum of the squared cross cumulants of y_i is also minimized. But JADE is restricted to small dimensions, mostly due to the computational complexity of the explicit tensor EVD. Its statistical properties are inferior to methods using likelihood or nonpolynomial

cumulants [22]. However, with low dimensional data, JADE is a competitive alternative to the most popular FastICA algorithms described in the next section.

A similar approach that uses the EVD is the fourth-order blind identification (FOBI) method [7], which is simpler and which deals with the EVD of the weighted correlation matrix. It is of reasonable complexity and is probably the most efficient of all the ICA methods. However, it fails to separate the sources when they have identical kurtosis. Other approaches include maximization of squared cumulants [23], and fourth-order cumulant based methods as described in [24,25].

5.7 Fast ICA Algorithm

One of the most popular solutions for linear ICA/BSS problem is Fast ICA [26], owing to its simplicity and fast convergence. The basic algorithm involves the preprocessing and a fixed-point iteration scheme for one unit.

5.7.1 Fixed-point Iteration for One Unit

The fast ICA algorithm for one unit estimates one row of the demixing matrix W as a vector w^T, which is an extremum of contrast functions. FastICA [19, 26] is an iterative fixed point algorithm, derived from a general objective or a contrast function. Assume x is the whitened data vector and w^T is one of the rows of the rotation/separating matrix W. Estimation of w proceeds iteratively with the following steps, until a convergence, as stated below, is achieved.

1) Choose an initial random vector w of unit norm.
2) $w \leftarrow E\{zg\,(w^T z)\} - Eg\,'(^T w)$ (40)

where $g_1(y) = y^3$ (derivative of kurtosis),

$g_2(y) = tanh(ay),\ 1 \le a \le 2$

and $g'(y)$ are the corresponding derivatives.

3) $w \leftarrow w\ /\ \|\ w\ \|$ where $\|w\|$ is the norm of w.
4) if $w_{old} - w_{new}$

$\le \varepsilon$ is not satisfied, then go back to step 2, where ε is a -4 convergence parameter (~10) and w_{old} is the value of w before its replacement by the newly calculated value w_{new}.

5.7.2 Fixed-point Iteration for Several Units

The independent components (ICs) can be estimated one by one, using the deflationary approach, or they can be estimated simultaneously, using the symmetric approach. In the deflationary approach, it must be ensured that the rows w_j of the separating matrix W are orthogonal. This can be done after every iteration step, by subtracting from the current estimate w_p the projections of all previously estimated p−1 vectors, before normalization.

$$W_p \leftarrow W_p - \sum_{j=1}^{p-1} (W_p^T W_j) W_j \tag{41}$$

In the symmetric approach, the iteration step is computed for all w_p and after the matrix W is orthogonalized, as -

$$W \leftarrow (WW^T)^{-1/2} W \tag{42}$$

The convergence properties of the FastICA algorithm are discussed in [26, 27]. The asymptotic convergence of the algorithm is at least quadratic and usually cubic when the ICA model (1) holds. This rate is much faster than that of gradient-based optimization algorithms. With a kurtosis-based contrast function, FastICA can be shown to converge globally to the independent components [19].

5.8 Algebraic ICA Algorithm

An algebraic solution to ICA is proposed by Taro Yamaguchi *et al.* in [28]. This is a noniterative algorithm but becomes extremely complex to compute when the number of sources goes more than two. For two sources separation, it works very fast. Two observed signals x_1 and x_2 are given by linear mixture of two independent original signals s_1 and s_2 as -

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & \alpha \\ \beta & 1 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \tag{43}$$

where α and β are unknown mixing rates.

The algebraic solution to α and β are given by -

$$\beta = \frac{\alpha C_2 - C_3}{\alpha C_3 - C_1} \tag{44}$$

$$(C_2C_{10} - C_{11}C_3)\alpha^4 + (3C_9C_3 - 3C_8C_2 - C_3C_{10} + C_1C_{11})\alpha^3$$
$$+ (3C_6C_2 + 3C_8C_3 - 3C_9C_1 - 3C_7C_3)\alpha^2$$
$$+ (C_5C_3 + 3C_7C_1 - 3C_6C_3 - 3C_2C_4)\alpha$$
$$+ (C_3C_4 - C_1C_5) = 0 \tag{45}$$

where

$C_1 = E[x_1^2] - \{E[x_1]\}^2$

$$C_2 = E[x_2^2] - \{E[x_2]\}^2$$

$$C_3 = E[x_1 x_2] - E[x_1]E[x_2]$$

$$C_4 = E[x_1^4] - E[x_1^3]E[x_1]$$

$$C_5 = E[x_1^3 x_2] - E[x_1^3]E[x_2]$$

$$C_6 = E[x_1^3 x_2] - E[x_1^2 x_2]E[x_1]$$

$$C_7 = E[x_1^2 x_2^2] - E[x_1^2 x_2]E[x_2]$$

$$C_8 = E[x_1^2 x_2^2] - E[x_1 x_2^2]E[x_1]$$

$$C_9 = E[x_1 x_2^3] - E[x_1 x_2^2]E[x_2]$$

$$C_{10} = E[x_1 x_2^3] - E[x_1]E[x_2^3]$$

$$C_{11} = E[x_2^4] - E[x_2^3]E[x_2] \quad (46)$$

where E[.] denotes the expectation operation.

α and β are obtained by solving the equations (44,45,46) with the Ferrari method. Excluding the solutions having non-zero imaginary parts and negative sizes, the proper solution is selected. Original independent signals are computed from equation (43) by solving value of α and β.

5.9 Evolutionary ICA Algorithm

Evolutionary computation techniques are very popular population search based optimization methods. Genetic Algorithms, Swarm intelligence are the most used evolutionary computation based optimization techniques. Through the evolutionary mechanism, genetic algorithms (GA) can search for the optimal separating matrix that minimizes the dependence. Instead of updating the matrix by a fixed formula, GA transforms a population of individuals into a new population using genetic operators, based on fitness function. However, the success of GA relies on the definition of fitness function. The population based search methods like GA converge to a global optimum, unlike the case of gradient based methods which get trapped in local optima. GA has been used for nonlinear blind source separation in [29,30] and for noise separation from electrocardiogram signals in [31]. Particle swarm optimization (PSO) has been used in ICA technique in [32] Currently, several biologically motivated optimization algorithms are also being used in ICA method. However, the price paid by evolutionary computation based ICA techniques is the heavy computational complexity of the methods. But with the advent of highly parallel processors, these methods provide competitive solutions to the problems.

5.10 Kernel ICA Algorithm

Kernel ICA [33] is a class of algorithms for independent component analysis (ICA), which use contrast functions based on canonical correlations in a reproducing kernel Hilbert space.

The ICA problem is based not on a single nonlinear function, but on an entire function space of candidate nonlinearities. In particular, the algorithm works with the functions in a reproducing kernel Hilbert space and makes use of the 'kernel trick' to search over this space efficiently. The use of a function space makes it possible to adapt to a variety of sources and thus makes these algorithms more robust to varying source distributions.

A contrast function is defined in terms of a direct measure of the dependence of a set of random variables. Considering the case of two univariate random variables x and x, for simplicity, and letting F be a vector space of functions from R to R, the F12 ρ correlation ^{F} is defined as the maximal correlation between the random variables $f(x_1)$ and $f(x_2)$, where $f_1$and $f_2$range over

$$\rho_F = \max_{f1 \Omega F} corr(f_1(x_1), f_2(x_2)) \quad (47)$$

If the variables x and x are independent, then the F-correlation is equal to zero. Also, if 12 the set F is large enough, the converse is true. Hence, the basic idea of Kernel ICA is first to map the input space into a feature space via a nonlinear map and then to extract the independent components from multivariate data. First, it estimates the dominant ICs and the directions using PCA and then it performs conventional ICA to update the dominant ICs while maintaining the variance.

The performance of Kernel ICA is robust, with respect to the source distributions. The Kernel ICA algorithms are particularly insensitive to asymmetry of the probability density function, when compared to the other algorithms. It is also reported [33] to yield smaller Amari error than other ICA algorithms.

Independent component analysis (ICA) algorithms are known to have difficulties when the sources are nearly Gaussian. The performance of all algorithms degrades as the kurtosis approaches zero, but the Kernel ICA algorithms are more robust to near-Gaussianity than other algorithms. The Kernel ICA methods are significantly more robust to outliers than the other ICA algorithms. However, they are slower than other algorithms.

5.11 Some Extensions to ICA Algorithm

5.11.1 Noisy ICA Algorithm

The estimation of the noiseless model seems to be a challenging task in itself, and, therefore, the noise is usually neglected, in order to obtain tractable and simple results. Moreover, it may be unrealistic in many cases to assume that the data could be divided into signals and noise in any meaningful way. Perhaps the most promising approach to noisy ICA is given by bias removal techniques. This means that noise free ICA methods are modified, so that the bias due to noise is removed or at least partially removed. In [34], bias reduction is performed by modifying the natural gradient ascent for likelihood. The new concept of Gaussian moments is introduced in [35], to derive one-unit contrast functions and to obtain a version of the fast ICA algorithm that has no asymptotic bias i.e. it is consistent even in the presence of noise. These techniques can even be used in large dimensions. In [36], J. Cao *et al.* have proposed a robust approach for independent component analysis (ICA) of signals that observations are contaminated with high-level additive noise and/or outliers.

5.11.2 Complex ICA Algorithm

Separation of complex valued signals is a frequently arising problem in signal processing. For example, separation of convolutively mixed source signals involves computations on complex valued signals. The FastICA algorithm can be extended to complex valued signals. In [37], it is assumed that the original, complex valued source signals are mutually statistically independent, and the problem is solved by the independent component analysis (ICA) model.

5.11.3 Nonlinear ICA Algorithm

In most of the practical cases, the linear mixtures pass through a certain type of nonlinearity, before being actually observed. Most often, the observing sensor introduces the nonlinearity by itself. So ICA must perform the separation from these observed nonlinear mixtures. The case of ICA for post nonlinear mixtures has been an area of interest for researchers [29,30].

6. Ambiguities of ICA

6.1 Permutation Ambiguity

The order of independent components cannot be determined. The linear noise free version of the ICA model can be represented as

$$X = \sum_{i=1}^{N} a_i s_i = AS \tag{48}$$

Both A and S being unknown, the order of the terms can be changed freely in the above equation and any of the independent components can be called the first one. This implies that the correspondence between a physical signal and the estimated independent component is not one-to-one. This indeterminacy is particularly severe in many applications, where identification of the estimated components is of very high importance. Formally, this means that the following relation between the mixing matrix A and the separation matrix B holds.

$$AB = P \tag{49}$$

where P is a permutation matrix.

6.2 Scaling Ambiguity

The energy of the independent components cannot be determined. Since both A and S are unknown, the effect of multiplication of one of the source estimates with a scalar constant k is canceled by dividing its corresponding column in the mixing matrix by k. This indeterminacy can be solved by ensuring that the random variables have unit variance i.e.,

$$E\{s_i^2\} = 1 \tag{50}$$

This still leaves the ambiguity of sign. While this is insignificant in certain applications, care has to be taken in applications where sign plays a crucial role.

7. Applications of ICA

Independent component analysis (ICA) being a blind statistical signal processing, the technique finds application in many emerging new application areas such as blind separation of mixed voices or images [38,39], analysis of several types of data [5], feature extraction [12], speech and image recognition [40,17], data communication [41], sensor signal processing [42,14], system identification [43,44], biomedical signal processing [45,46,13,29] and several others [30,47].

7.1 Biomedical Signal Processing

Magnetoencephalography (MEG) is a noninvasive technique by which the activity or the cortical neurons can be measured with very good temporal resolution and moderate spatial resolution. When using a MEG record, as a research or clinical tool, the investigator may face a problem of extracting the essential features of the neuromagnetic signals in the presence of artifacts. The amplitude of the disturbances may be higher than that of the brain signals and the artifacts may resemble pathological signals in shape. In [48], a new method to separate brain activity from artifacts using ICA has been introduced.

7.2 Telecommunications

Finally, another emerging application area of great potential is telecommunications. An example of a real-world communications application where blind separation techniques are useful is the separation of the user's own signal from the interfering other users' signals in CDMA (Code-Division Multiple Access) mobile communications [49]. This problem is semi-blind, in the sense that certain additional prior information is available on the CDMA data model. But the number of parameters to be estimated is often so high that suitable blind source separation techniques, taking into account the available prior knowledge, provide a clear performance improvement over more traditional estimation techniques.

7.3 Revealing Hidden Factors in Financial Data

It is a tempting alternative to try ICA on financial data. There are many situations in that domain of finance in which parallel time series are available, such as currency exchange rates or daily returns of stocks, which may have some common underlying factors. Independent component analysis (ICA) might reveal some driving mechanisms that may otherwise remain hidden. In a recent study of a stock portfolio [50] it was found that ICA is a complementary tool to PCA, allowing the underlying structure of the data to be more readily observed.

7.4 Natural Image Denoising

Bell and Sejnowski proposed a method to extract features from natural scenes by assuming a linear image synthesis model [51]. In such a model, each patch of an image is a linear combination of several underlying basis functions. A set of digitized natural images were used. The vector of pixel gray levels in an image window is denoted by **x**. Note that, multi-valued time series or images changing with time are not considered here; instead the elements of x are indexed by the location in the image window or patch. The sample windows were taken at random locations. The 2-D structure of the windows is of no significance here; row by row scanning was used to turn a square image window into a vector of pixel values. Each window corresponds to one of the columns *ai* of the mixing matrix A. Thus, an observed image window is a superposition of these windows with independent coefficients [51].

Now, suppose a noisy image model holds:

$$z = x + n \tag{51}$$

where n is uncorrelated noise, with elements indexed in the image window in the same way as x, and z is the measured image window corrupted with noise. Let us further assume that **n** is Gaussian and **x** is nongaussian. There are many ways to clean the noise. One example is to make a transformation to spatial frequency space by discrete fourier transform (DFT), do low-pass filtering, and to return to the image space by inverse discrete fourier transform (IDFT). This is not very efficient, however. A better method is the recently introduced Wavelet Shrinkage method [52], in which a transform based on wavelets is used, or methods based on median filtering are used. However, it may be noted that none of these methods explicitly takes advantage of the image statistics.

7.5 Feature Extraction

Independent component analysis (ICA) is successfully used for face recognition and lip reading. The goal in face recognition is to train a system that can recognize and classify familiar faces, given a different image of the trained face. The test images may show the faces in a different pose or under different lighting conditions. Traditional methods for face recognition have employed PCA-like methods. Bartlett and Sejnowski [53] compare the face recognition performance of PCA and ICA for two different tasks: (1) different pose and (2) different lighting condition. They show that for both tasks, ICA outperforms PCA. The method is roughly as follows: The rows of the face images constitute the data matrix x. Performing ICA, a transformation W is learned so that u (u = Wx) represents the independent face images. The nearest neighbor classification is performed on the coefficients of u.

7.6 Nonlinear Process Monitoring

The production processes of chemical, pharmaceutical and biological products being nonlinear involve intricate methods of monitoring. Zhang and Qin in [54] develop a process monitoring method based on multiway kernel independent component analysis, which extracts some dominant independent components that capture nonlinearity from normal operating process data and combine them with statistical process monitoring techniques. They apply the method to fault detection in a fermentation process. However, there are certain drawbacks of original KPCA and KICA, which are as follows: the data mapped into feature space become redundant; linear data introduce errors while the kernel trick is used; computation time increases with the number of samples. In [55] Zhang and Qin improve KPCA and KICA for nonlinear fault detection and statistical analysis.

A novel technique is proposed in [56], which combines the advantage of both kernel principal component

analysis (KPCA) and Kernel ICA (KICA) to develop a nonlinear dynamic approach to detect fault online, compared to other nonlinear approaches.

7.7 Sensor Signal Processing

A sensor network is a very recent, widely applicable and challenging field of research. As the size and cost of sensors decrease, sensor networks are increasingly becoming an attractive method to collect information in a given area. Multi-sensor data often presents complementary information about the region surveyed and data fusion provides an effective method to enable comparison, interpretation and analysis of such data. Image and video fusion is a sub area of the more general topic of data fusion, dealing with image and video data. Cvejic *et al* [42]. have applied independent component analysis for improving the fusion of multimodal surveillance images in sensor networks. Independent component analysis (ICA) is also used for robust automatic speech recognition [57].

8. Conclusions

In this chapter, the basic principle behind the independent component analysis (ICA) technique is discussed. The contrast functions for different routes to independence are clearly depicted. Different existing algorithms for ICA are briefly illustrated and are critically examined, with special reference to their algorithmic properties. The ambiguities present in these algorithms are also presented. Finally, the application domains of this novel technique are presented. Some of the futuristic works on ICA technique, which need further investigation, are development of nonlinear ICA algorithms, design of low complexity ICA algorithms and use of evolutionary computing optimization tools for developing ICA and, finally, alleviation of permutation and scaling ambiguities existing in present ICA.

References

1. P. Comon, "Independent Component Analysis-A new concept?" *Signal Processing*, vol. 36, pp. 287-314, 1994.
2. J.F.Cardoso, "Blind Signal Separation: Statistical Principles", *Proc. of IEEE*, vol. 9, no. 10, pp. 2009-2025, 1998.
3. S.Z.Li, *et al*., "Learning Multiview Face Subspaces and Facial Pose Estimation using Independent Component Analysis," *IEEE Trans. Image Processing*, vol. 14, no. 6, pp. 705-712, June 2005.
4. Capizzi G., Coco S. and Laudani A., "A New Tool for the Identification and Localization of Electromagnetic Sources by Using Independent Component Analysis", *IEEE Trans. On Magnetics*, vol. 43, Issue-4, pp. 1625-1628, Apr. 2007.
5. C.Jutten and J.Herault, "Blind separation of sources, part I: An adaptive algorithm based on neuromimetric architecture," *Signal Processing*, 24:1-10, 1991.
6. J.F. Cardoso and A. Souloumiac. Blind beamforming for non Gaussian signals. *IEE Proceedings-F*, 140(6), pp. 362-370, 1993.
7. J.F.Cardoso "Source separation using higher order moments", *Proc. of the IEEE Int. Conf on Acoustics, Speech and Signal Processing (ICASSP 1989)*, pp. 2109-2112, Glasgow, UK, May 1989.
8. K. Fukunaga and D. R. Olsen. "An algorithm for finding intrinsic dimensionality of data" *IEEE Transactions on Computers*, 202:176-183, 1971.
9. D.T.Pham, P.Garrat and C.Jutten, "Separation of a mixture of independent sources through a maximum likelihood approach", *Proc. EUSIPCO*, pp. 771-774, 1992.
10. E.Oja, H.Ogawa, and J.Wangviwattana, "Learning in nonlinear constrained Hebbian networks", T.Kohonen, *et al*., editor, Artificial Neural Networks, *Proc. ICANN'91*, pp. 385-390, Espoo, Finland, 1991.
11. E.Oja, "Nonlinear PCA criterion and maximum likelihood in independent component analysis", *Proc. Int. Workshop on Independent Component Analysis and Signal separation (ICA'99)*, pp. 143-148, Aussois, France, 1999.
12. N.Delfosse and P.Loubaton, "Adaptive blind separation of independent sources: a deflation approach," *Signal Processing*, 45:59-83, 1995.
13. A.Cichocki "Robust Neural Networks with on line biasing for blind identification and blind source separation", *IEEE Trans. On Circuits and Systems*, vol. 43, no. 11, pp. 894-906, 1996.
14. J.F.Cardoso and B.H.Laheld, "Equivariant adaptive source separation", *IEEE Trans. On Signal Processing*, vol. 44, no. 12, pp. 3017-3030, 1996.
15. B. Laheld and J.F.Cardoso, "Adaptive source separation with uniform performance", *Proc. of EUSIPCO*, pp. 183-186, Edinburgh, 1994.
16. A.J.Bell and T.J.Sejnowski, "An information-maximization approach to blind separation and blind decovolution", *Neural Computation*, 7, pp. 1129-1159, 1995.
17. J.F.Cardoso, "Infomax and maximum likelihood for sources separation," *IEEE Letters on Signal Processing*, 4:112-114, 1997.
18. H. Hotelling. "Analysis of a complex of statistical variables into principal components", *Journal of Educational Psychology*, 24:417-441, 498-520, 1993.
19. A. Hyvarinen and E. Oja "A fast fixed-point algorithm for independent component analysis", *Neural Computation*, 9, pp. 1483-1492, 1997.
20. A.Hyverinen and E.Oja, "Independent component analysis by general nonlinear Hebbian like learning rates", *Signal Processing*, vol. 64, no. 3, pp. 301-303, 1998.
21. T.Kohonen, "*Self Organizing Maps*", Springer-Verlag, Berlin, Heidelberg, New York, 1995.
22. A. Hyverinen., J.Kahrunen. and E. Oja, " *Independent Component Analysis*", John Wiley and Sons, 2001.
23. F. Herrmann and A. K. Nandi, "Maximisation of squared cumulants for blind source separation" *Electronics Letters*, 36(19), pp. 1664-1665, 1996.
24. A. K. Nandi and F. Herrmann, "Fourth-order cumulant based estimator for independent component analysis", *Electronic Letters*, 37(7), pp. 469-470, 2001.
25. A. K. Nandi and V. Zarzoso, "Fourth-order cumulant based blind source separation", *IEEE Signal Processing Letters*, 3(12), pp. 312-314, 1996.
26. A.Hyverinen, "Fast and Robust Fixed-point Algorithm for Independent Component Analysis", *IEEE Trans. on Neural Networks*, vol. 10, no. 3, pp. 626-634, 1999a.
27. E. Oja. "Convergence of the symmetrical FastICA algorithm", *Proc. of the 9th Int. Conf. on Neural Information Processing (ICONIP 2002)*, Singapore, Nov. 2002.
28. T.Yamaguchi, I. Kuzuyoshi., "An Algebraic Solution to Independent Component Analysis", *Optics Communications, Elsevier Science*, 178, pp. 59-64, 2000.
29. Y.Tan and J.Wang, "Nonlinear Blind Source Separation Using Higher Order Statistics and a Genetic Algorithm," *IEEE Trans. On Evolutionary Computation*, vol. 5, no. 6, pp. 600-611, Dec. 2001.
30. F.Rojas, C.G.Puntonet, M.Rodriguez-Alvarez, I.Rojas, and R.Martin-Clemente, "Blind Source separation in post-nonlinear mixtures using competitive learning, simulated annealing and a genetic algorithm," *IEEE Trans. On Systems, Man and Cybernatics -Part C: Applications and Reviews*, vol. 34, no. 4, pp. 407-416, Nov. 2004.

31. R.Palaniappan and C.N.Gupta, "Genetic Algorithm based independent component analysis to separate noise from Electrocardiogram signals," *Proc. IEEE*, 2006.
32. D. J. Krusienski and W. K. Jenkins, "Nonparametric Density Estimation Based Independent Component Analysis via Particle Swarm Optimization", *Proc. IEEE*, pp. IV - 357-IV - 360, 2005.
33. Francis R. Bach and Michael I. Jordan, "Kernel Independent Component Analysis", *Journal of Machine Learning Research* 3, pp. 1-48, 2002.
34. S.C.Douglas, A.Cichocki and S. Amari, "A bias removal technique for blind source separation with noisy measurements", *Electronics Letters*, 34, pp. 1379-1380, 1998.
35. A.Hyvarinen, "Fast Independent Component analysis with noisy data using Gaussian moments", *Proc. Int. Symp. On Circuits and Systems*, Orlando, Florida, pp. 57-61, 1999.
36. J. Cao, N. Murata, S. Amari, A. Cichocki, and T. Takeda "A Robust Approach to Independent Component Analysis of Signals With High-Level Noise Measurements", *IEEE Trans. on Neural Networks*, vol. 14, no. 3, pp. 631-645,May 2003.
37. E.Bingham and A. Hyvarinen, "A Fast fixed-point algorithm for independent component analysis of complex valued signals", *International Journal of Neural Systems*, vol. 10, no. 1, pp. 1-8, Feb. 2000.
38. I.Dagher, R.Nachar, "Face Recognition using IPCA-ICA Algorithm," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 6, pp. 996-1000, June 2006.
39. Keun-Chang Kwak, Pedrycz,W, "Face Recognition using Enhanced Independent Component Analysis Approach", *IEEE Trans. on Neural Networks*, vol. 18, no. 4, pp. 530-541, Mar. 2007.
40. Keun-Chang Kwak and W.Pedrycz, "Face Recognition using an Enhanced Independent Component Analysis Approach," *IEEE Trans. On Neural Networks*, vol. 18, Issue-2, pp. 1625-1628, Apr. 2007.
41. E.Oja, H.Ogawa, and J. Wangviwattana, "Learning in nonlinear constrained Hebbian networks," In T.Kohonen, *et al.*,editor, Artificial Neural Networks,Proc.ICANN'91, pp. 385-390, Espoo, Finland, 1991. North-Holland, Amsterdam.
42. Cvejic N., Bull D. and Canagarajah N., "Improving Fusion of Surveilliance Images in Sensor Networks using Independent Component Analysis", *IEEE Trans. On Consumer Electronics*, vol. 53, Issue 3, pp. 1029-1035, Aug. 2007.
43. Jun-Mei Yang and Sakai, H., "A New Adaptive Filter Algorithm for System Identification using Independent Component Analysis", *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2007, vol. 3, pp. III-1341-III-1344, Apr. 2007
44. Ou Shifeng, Zhao Xiaohui and Gao Ying, "Linear System Identification Employing Independent Component Analysis", *IEEE Int. Conf. on Automation and Logistics*, 2007, pp. 1336-1340, Aug. 2007.
45. Van Dun B., Wouters J. and Moonen M., "Improving Auditory Steady-State Response Detection Using Independent Component Analysis on Multichannel EEG Data", *IEEE Trans. On Biomedical Engineering*, vol. 54, Issue 7, pp. 1220-1230, July 2007.
46. Waldert S "Real-Time fetal heart Monitoring in Biomagnetic Measurements Using Adaptive Real -Time ICA", *IEEE Trans. On Biomedical Engineering*, vol. 54, Issue 107, pp. 1864-1874, Oct. 2007.
47. Vitria J, Bressan M. and Radeva P., "Bayesian Classification of Cork Stoppers Using Class-Conditional Independent Component Analysis", *IEEE Trans. On Systems, Man and Cybernatics -Part C: Applications and Reviews*, vol. 37, Issue.1, pp. 32-38, Jan. 2007.
48. R.Vigário,, J. Särelä,, and Oja, E., "Independent component analysis in wave decomposition of auditory evoked fields", *Proc. Int. Conf. on Artificial Neural Networks*, pp. 287-292, Skövde, Sweden, 1998.
49. T. Ristaniemi, and J. Joutsensalo,, "On the performance of blind source separation in CDMA downlink", *Proc. Int. Workshop on Independent Component Analysis and Signal Separation*, pp. 437-441, Aussois, France, 1999.
50. A. D.Back, and A.S.Weigend,, "A first application of independent component analysis to extracting structure from stock returns", *Int. J. on Neural Systems*, 8(4), pp. 473-484, 1997.
51. A.J.Bell and T.J.Sejniwski, "The 'Independent Components' of natural scenes are edge filters", *Vision Research*, 37(23), pp. 3327-3338, 1997.
52. D.L. Donoho, I.M.Johnstone, G.Kerkyacharian, and D.Picard, "Wavelet shrinkage: asymptopia?" *Journal of the Royal Statistical Society, Ser. B*, 57, pp. 301- 337, 1995.
53. M.Bartlett and T.J.Sejniwski, "Viewpoint invariant face recognition using independent component analysis and attractor networks", *Advances in Neural Information Processing Systems* 9, pp. 817-823, MIT Press, 1997.
54. Y.Zhang and S.J.Qin, "Fault Detection of Nonlinear Processes Using Multiway Kernel Independent Component Analysis", *Industrial Engineering Chemistry Research*, 46, pp. 7780-7787, 2007.
55. Y.Zhang and S.J.Qin, "Improved Nonlinear Fault Detection Technique and Statistical Analysis", *American Institute of Chemical Engineers Journal*, vol. 54, no. 12, pp. 3207-3220, Dec 2008.
56. Y.Zhang, "Enhanced Statistical Analysis of Nonlinear Processes Using KPCA, KICA and SVM", *Chemical Engineering Science*, 2008,(Press).
57. L.Potanutis "Independent Component Analysis applied to Feature Extraction for Robust Automatic Speech Recognition," *Electronics Lett.*, vol. 36, no. 23, pp. 1977-1978, Nov. 2000.

AUTHORS

Debiprasad Priyabrata Acharya is presently working as a faculty in the Dept. of ECE at NIT Rourkela. He did his M.Sc. (Physics), M.E.(Microelectronics) and Ph.D. (VLSI Signal Processing) from Indian Institute of Technology, Kharagpur, India, Birla Institute of Technology and Science, Pilani, India and National Institute of Technology, Rourkela, India in 1996, 2002 and 2008 respectively. He has worked as a Research Engineer at Center for Development of Telematics, Bangalore, Technical Consultant to Wipro Technologies, Bangalore and a Specialist at Tata Elxsi Ltd, Bangalore since 2002 till 2007. His present research interests include VLSI Design, Wireless Communications and Signal Processing.

E-mail: d_p_acharya@rediffmail.com

Prof. Ganapati Panda is presently working as the Professor & Head in the Dept. of E&CE at NIT, Rourkela. Prior to this he was the Director of NIT, Jamshedpur. He did his Post Doc. from University of Edinburgh, UK and Ph. D. from IIT, Kharagpur, India.

He already guided 16 Ph. Ds scholars and published more than 232 research papers in various International and National referred journals and conferences. He is a Fellow of National Academy of Engineering (FNAE) and Fellow of National Academy of Science (FNASc.). He received Samanta Chandra Sekhar award for his significant research contribution from Govt. of Orissa in 1993. His present research interests are DSP, Digital Communication, Intelligent Instrumentation, Soft Computing and Sensor Networks.

E-mail: ganapati.panda@gmail.com

Paper No TR 51_08; Copyright © 2008 by the IETE