

# An Overview of Independent Component Analysis and Its Applications

Ganesh R. Naik and Dinesh K Kumar  
 School of Electrical and Computer Engineering  
 RMIT University, Australia  
 E-mail: ganesh.naik@rmit.edu.au

## Overview paper

**Keywords:** independent component analysis, blind source separation, non-gaussianity, multi run ICA, overcomplete ICA, undercomplete ICA

**Received:** July 3, 2009

*Independent Component Analysis (ICA), a computationally efficient blind source separation technique, has been an area of interest for researchers for many practical applications in various fields of science and engineering. This paper attempts to cover the fundamental concepts involved in ICA techniques and review its applications. A thorough discussion of the applications and ambiguities problems of ICA has been carried out. Different ICA methods and their applications in various disciplines of science and engineering have been reviewed. In this paper, we present ICA methods from the basics to their potential applications to serve as a comprehensive single source for an inquisitive researcher to carry out research in this field.*

*Povzetek: Podan je pregled tehnike ICA (Independent Component Analysis).*

## 1 Introduction

The problem of source separation is an inductive inference problem. There is not enough information to deduce the solution, so one must use any available information to infer the most probable solution. The aim is to process these observations in such a way that the original source signals are extracted by the adaptive system. The problem of separating and estimating the original source waveforms from the sensor array, without knowing the transmission channel characteristics and the source can be briefly expressed as problems related to BSS. In BSS the word blind refers to the fact that we do not know how the signals were mixed or how they were generated. As such, the separation is in principle impossible. Allowing some relatively indirect and general constraints, we however still hold the term BSS valid, and separate under these conditions.

There appears to be something magical about blind source separation; we are estimating the original source signals without knowing the parameters of mixing and/or filtering processes. It is difficult to imagine that one can estimate this at all. In fact, without some a priori knowledge, it is not possible to uniquely estimate the original source signals. However, one can usually estimate them up to certain indeterminacies. In mathematical terms, these indeterminacies and ambiguities can be expressed as arbitrary scaling, permutation and delay of estimated source signals [1]. These indeterminacies preserve, however, the waveforms of the original sources. Although these indeterminacies seem to be rather severe limitations, in a great number of applications these limitations are not essential, since the most relevant information about the source signals

is contained in the temporal waveforms or time-frequency patterns of the source signals and usually not in their amplitudes or the order in which they are arranged in the output of the system. However, for some applications especially biomedical signal models such as sEMG signals, there is no guarantee that the estimated or extracted signals have exactly the same waveforms as the source signals.

Independent component analysis (ICA) is one of the most widely used BSS techniques for revealing hidden factors that underlie sets of random variables, measurements, or signals. ICA is essentially a method for extracting individual signals from mixtures. Its power resides in the physical assumptions that the different physical processes generate unrelated signals. The simple and generic nature of this assumption allows ICA to be successfully applied in diverse range of research fields.

In this paper, we first set the scene of the blind source separation problem. Then, Independent Component Analysis is introduced as a widely used technique for solving the blind source separation problem. A general description of the approach to achieving separation via ICA and the underlying assumptions of the ICA framework and important ambiguities that are inherent to ICA are discussed in section 3. A description of specific details of different ICA methods are given in Sections 4, and the paper concludes with applications of BSS and ICA methods.

## 2 Blind source separation (BSS)

Consider a situation in which we have a number of sources emitting signals which are interfering with one another. Fa-

miliar situations in which this occurs are a crowded room with many people speaking at the same time, interfering electromagnetic waves from mobile phones or crosstalk from brain waves originating from different areas of the brain. In each of these situations the mixed signals are often incomprehensible and it is of interest to separate the individual signals. This is the goal of Blind Source Separation. A classic problem in BSS is the cocktail party problem. The objective is to sample a mixture of spoken voices, with a given number of microphones - the observations, and then separate each voice into a separate speaker channel - the sources. The BSS is unsupervised and thought of as a black box method. In this we encounter many problems, e.g. time delay between microphones, echo, amplitude difference, voice order in speaker and underdetermined mixture signal.

Herault and Jutten [2] proposed that, in a artificial neural network like architecture the separation could be done by reducing redundancy between signals. This approach initially lead to what is known as independent component analysis today. The fundamental research involved only a handful of researchers up until 1995. It was not until then, when Bell and Sejnowski [3] published a relatively simple approach to the problem named infomax, that many became aware of the potential of ICA. Since then a whole community has evolved around ICA, centralized around some large research groups and its own ongoing conference, International Conference on independent component analysis and blind signal separation. ICA is used today in many different applications, e.g. medical signal analysis, sound separation, image processing, dimension reduction, coding and text analysis [4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14].

In ICA the general idea is to separate the signals, assuming that the original underlying source signals are mutually independently distributed. Due to the field's relatively young age, the distinction between BSS and ICA is not fully clear. When regarding ICA, the basic framework for most researchers has been to assume that the mixing is instantaneous and linear, as in infomax. ICA is often described as an extension to PCA, that uncorrelates the signals for higher order moments and produces a non-orthogonal basis. More complex models assume for example, noisy mixtures, [15, 16], nontrivial source distributions, [17, 18], convolutive mixtures [19, 20, 21], time dependency, underdetermined sources [22, 23], mixture and classification of independent component [4, 24]. A general introduction and overview can be found in [25].

### 3 Independent component analysis

Independent Component Analysis (ICA) is a statistical technique, perhaps the most widely used, for solving the blind source separation problem [25, 26]. In this section, we present the basic Independent Component Analysis model and show under which conditions its parameters can be estimated.

#### 3.1 ICA model

The general model for ICA is that the sources are generated through a linear basis transformation, where additive noise can be present. Suppose we have  $N$  statistically independent signals,  $s_i(t), i = 1, \dots, N$ . We assume that the sources themselves cannot be directly observed and that each signal,  $s_i(t)$ , is a realization of some fixed probability distribution at each time point  $t$ . Also, suppose we observe these signals using  $N$  sensors, then we obtain a set of  $N$  observation signals  $x_i(t), i = 1, \dots, N$  that are mixtures of the sources. A fundamental aspect of the mixing process is that the sensors must be spatially separated (e.g. microphones that are spatially distributed around a room) so that each sensor records a different mixture of the sources. With this spatial separation assumption in mind, we can model the mixing process with matrix multiplication as follows:

$$x(t) = As(t) \quad (1)$$

where  $A$  is an unknown matrix called the *mixing matrix* and  $x(t), s(t)$  are the two vectors representing the observed signals and source signals respectively. Incidentally, the justification for the description of this signal processing technique as *blind* is that we have no information on the mixing matrix, or even on the sources themselves.

The objective is to recover the original signals,  $s_i(t)$ , from only the observed vector  $x_i(t)$ . We obtain estimates for the sources by first obtaining the "unmixing matrix"  $W$ , where,  $W = A^{-1}$ .

This enables an estimate,  $\hat{s}(t)$ , of the independent sources to be obtained:

$$\hat{s}(t) = Wx(t) \quad (2)$$

The diagram in Figure 1 illustrates both the mixing and unmixing process involved in BSS. The independent sources are mixed by the matrix  $A$  (which is unknown in this case). We seek to obtain a vector  $y$  that approximates  $s$  by estimating the unmixing matrix  $W$ . If the estimate of the unmixing matrix is accurate, we obtain a good approximation of the sources.

The above described ICA model is the simple model since it ignores all noise components and any time delay in the recordings.

#### 3.2 Independence

A key concept that constitutes the foundation of independent component analysis is statistical independence. To simplify the above discussion consider the case of two different random variables  $s_1$  and  $s_2$ . The random variable  $s_1$  is independent of  $s_2$ , if the information about the value of  $s_1$  does not provide any information about the value of  $s_2$ , and vice versa. Here  $s_1$  and  $s_2$  could be random signals originating from two different physical process that are not related to each other.

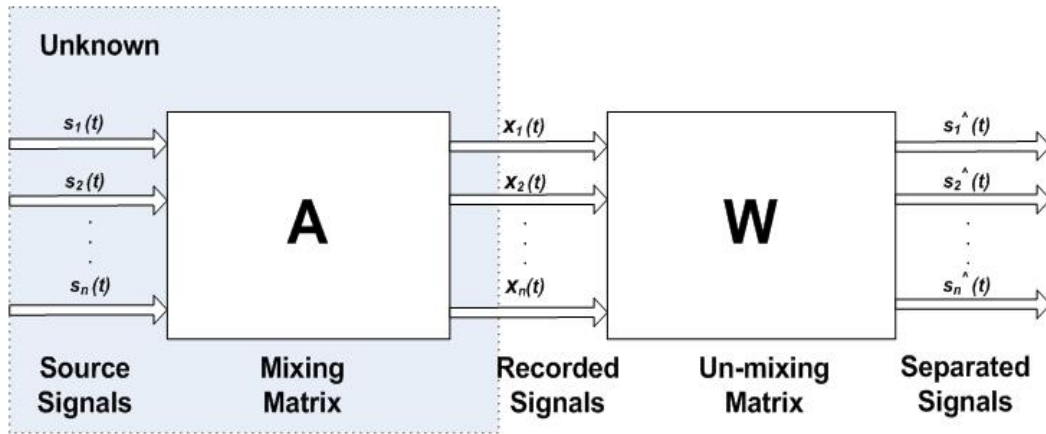


Figure 1: Blind source separation (BSS) block diagram.  $s(t)$  are the sources.  $x(t)$  are the recordings,  $\hat{s}(t)$  are the estimated sources  $A$  is mixing matrix and  $W$  is un-mixing matrix

**3.2.1 Independence definition**

Mathematically, statistical independence is defined in terms of probability density of the signals. Consider the joint probability density function (pdf) of  $s_1$  and  $s_2$  be  $p(s_1, s_2)$ . Let the marginal pdf of  $s_1$  and  $s_2$  be denoted by  $p_1(s_1)$  and  $p_2(s_2)$  respectively.  $s_1$  and  $s_2$  are said to be independent if and only if the joint pdf can be expressed as;

$$p_{s_1, s_2}(s_1, s_2) = p_1(s_1)p_2(s_2) \tag{3}$$

Similarly, independence could be defined by replacing the pdf by the respective cumulative distributive functions as;

$$E\{p(s_1)p(s_2)\} = E\{g_1(s_1)\}E\{g_2(s_2)\} \tag{4}$$

where  $E\{\cdot\}$  is the expectation operator. In the following section we use the above properties to explain the relationship between uncorrelated and independence.

**3.2.2 Uncorrelatedness and Independence**

Two random variables  $s_1$  and  $s_2$  are said to be uncorrelated if their covariance  $C(s_1, s_1)$  is zero.

$$\begin{aligned} C(s_1, s_2) &= E\{(s_1 - m_{s1})(s_2 - m_{s2})\} \\ &= E\{s_1s_2 - s_1m_{s2} - s_2m_{s1} + m_{s1}m_{s2}\} \\ &= E\{s_1s_2\} - E\{s_1\}E\{s_2\} \\ &= 0 \end{aligned} \tag{5}$$

where  $m_{s1}$  is the mean of the signal. Equation 4 and 5 are identical for independent variables taking  $g_1(s_1) = s_1$ . Hence independent variables are always uncorrelated. However the opposite is not always true. The above discussion proves that independence is stronger than uncorrelatedness and hence independence is used as the basic principle for ICA source estimation process. However uncorrelatedness is also important for computing the mixing matrix in ICA.

**3.2.3 Non-Gaussianity and Independence**

According to central limit theorem the distribution of a sum of independent signals with arbitrary distributions tends toward a Gaussian distribution under certain conditions. The sum of two independent signals usually has a distribution that is closer to Gaussian than distribution of the two original signals. Thus a gaussian signal can be considered as a liner combination of many independent signals. This furthermore elucidate that separation of independent signals from their mixtures can be accomplished by making the linear signal transformation as non-Gaussian as possible.

Non-Gaussianity is an important and essential principle in ICA estimation. To use non-Gaussianity in ICA estimation, there needs to be quantitative measure of non-Gaussianity of a signal. Before using any measures of non-Gaussianity, the signals should be normalised. Some of the commonly used measures are kurtosis and entropy measures, which are explained next.

**– Kurtosis**

Kurtosis is the classical method of measuring Non-Gaussianity. When data is preprocessed to have unit variance, kurtosis is equal to the fourth moment of the data.

The Kurtosis of signal ( $s$ ), denoted by  $kurt(s)$ , is defined by

$$kurt(s) = E\{s^4\} - 3(E\{s^2\})^2 \tag{6}$$

This is a basic definition of kurtosis using higher order (fourth order) cumulant, this simplification is based on the assumption that the signal has zero mean. To simplify things, we can further assume that ( $s$ ) has been normalised so that its variance is equal to one:  $E\{s^2\} = 1$ .

Hence equation 6 can be further simplified to

$$kurt(s) = E\{s^4\} - 3 \tag{7}$$

Equation 7 illustrates that kurtosis is a nomralised form of the fourth moment  $E\{s^4\} = 1$ . For Gaussian signal,

$E\{s^4\} = 3(E\{s^2\})^2$  and hence its kurtosis is zero. For most non-Gaussian signals, the kurtosis is nonzero. Kurtosis can be both positive or negative. Random variables that have positive kurtosis are called as *super Gaussian* or *platykurtotic*, and those with negative kurtosis are called as *sub Gaussian* or *leptokurtotic*. Non-Gaussianity is measured using the absolute value of kurtosis or the square of kurtosis.

Kurtosis has been widely used as measure of Non-Gaussianity in ICA and related fields because of its computational and theoretical and simplicity. Theoretically, it has a linearity property such that

$$kurt(s_1 \pm s_2) = kurt(s_1) \pm kurt(s_2) \quad (8)$$

and

$$kurt(\alpha s_1) = \alpha^4 kurt(s_1) \quad (9)$$

where  $\alpha$  is a constant. Computationally kurtosis can be calculated using the fourth moment of the sample data, by keeping the variance of the signal constant.

In an intuitive sense, kurtosis measured how "spikiness" of a distribution or the size of the tails. Kurtosis is extremely simple to calculate, however, it is very sensitive to outliers in the data set. Its values may be based on only a few values in the tails which means that its statistical significance is poor. Kurtosis is not robust enough for ICA. Hence a better measure of non-Gaussianity than kurtosis is required.

#### – Entropy

Entropy is a measure of the uniformity of the distribution of a bounded set of values, such that a complete uniformity corresponds to maximum entropy. From the information theory concept, entropy is considered as the measure of randomness of a signal. Entropy  $H$  of discrete-valued signal  $S$  is defined as

$$H(S) = -\sum P(S = a_i) \log P(S = a_i) \quad (10)$$

This definition of entropy can be generalised for a continuous-valued signal ( $s$ ), called differential entropy, and is defined as

$$H(S) = -\int p(s) \log p(s) ds \quad (11)$$

One fundamental result of information theory is that Gaussian signal has the largest entropy among the other signal distributions of unit variance. entropy will be small for signals that have distribution concerned on certain values or have pdf that is very "spiky". Hence, entropy can be used as a measure of non-Gaussianity.

In ICA estimation, it is often desired to have a measure of non-Gaussianity which is zero for Gaussian signal and nonzero for non-Gaussian signal for computational simplicity. Entropy is closely related to the code length of the random vector. A normalised version of entropy is given by a new measure called Negentropy  $J$  which is defined as

$$J(S) = H(s_{gauss}) - H(s) \quad (12)$$

where  $s_{gauss}$  is the Gaussian signal of the same covariance matrix as ( $s$ ). Equation 12 shows that Negentropy is always positive and is zero only if the signal is a pure gaussian signal. It is stable but difficult to calculate. Hence approximation must be used to estimate entropy values.

### 3.3 Mathematical Independence

Mathematical properties of matrices were investigated to check the linear dependency and independency of global matrices (Permutation matrix  $P$ )

#### 3.3.1 Rank of the matrix

Rank of the matrix will be less than the matrix size for linear dependency and rank will be size of matrix for linear independency, but this couldn't be assured yet due to noise in the signal. Hence determinant is the key factor for estimating number of sources.

#### 3.3.2 Determinant of the matrix

In real time applications Determinant value should be zero for linear independency and should be more than zero (close to 1) for linear independency [27].

### 3.4 ICA Assumptions and Ambiguities

ICA is distinguished from other approaches to source separation in that it requires relatively few assumptions on the sources and on the mixing process. The assumptions and of the signal properties and other conditions and the issues related to ambiguities are discussed below:

#### 3.4.1 ICA Assumptions

– *The sources being considered are statistically independent*

The first assumption is fundamental to ICA. As discussed in Section 3.2, statistical independence is the key feature that enables estimation of the independent components  $\hat{s}(t)$  from the observations  $x_i(t)$ .

– *The independent components have non-Gaussian distribution*

The second assumption is necessary because of the close link between Gaussianity and independence. It is impossible to separate Gaussian sources using the ICA framework described in Section 3.2 because the sum of two or more Gaussian random variables is itself Gaussian. That is, the sum of Gaussian sources is indistinguishable from a single Gaussian source in the ICA framework, and for this reason Gaussian sources are forbidden. This is not an overly restrictive assumption as in practice most sources of interest are non-Gaussian.

– *The mixing matrix is invertible*

The third assumption is straightforward. If the mixing matrix is not invertible then clearly the unmixing matrix we seek to estimate does not even exist.

If these three assumptions are satisfied, then it is possible to estimate the independent components modulo some trivial ambiguities (discussed in Section 3.4). It is clear that these assumptions are not particularly restrictive and as a result we need only very little information about the mixing process and about the sources themselves.

### 3.4.2 ICA Ambiguity

There are two inherent ambiguities in the ICA framework. These are (i) magnitude and scaling ambiguity and (ii) permutation ambiguity.

– *Magnitude and scaling ambiguity*

The true variance of the independent components cannot be determined. To explain, we can rewrite the mixing in equation 1 in the form

$$\begin{aligned} x &= As \\ &= \sum_{j=1}^N a_j s_j \end{aligned} \quad (13)$$

where  $a_j$  denotes the  $j$ th column of the mixing matrix  $A$ . Since both the coefficients  $a_j$  of the mixing matrix and the independent components  $s_j$  are unknown, we can transform Equation 13.

$$x = \sum_{j=1}^N (1/\alpha_j a_j)(\alpha_j s_j) \quad (14)$$

Fortunately, in most of the applications this ambiguity is insignificant. The natural solution for this is to use assumption that each source has unit variance:  $E\{s_j^2\} = 1$ . Furthermore, the signs of the of the sources cannot be determined too. This is generally not a serious problem because the sources can be multiplied by -1 without affecting the model and the estimation

– *Permutation ambiguity*

The order of the estimated independent components is unspecified. Formally, introducing a permutation matrix  $P$  and its inverse into the mixing process in Equation 1.

$$\begin{aligned} x &= AP^{-1}Ps \\ &= A' s \end{aligned} \quad (15)$$

Here the elements of  $P$   $s$  are the original sources, except in a different order, and  $A' = AP^{-1}$  is another unknown mixing matrix. Equation 15 is indistinguishable from Equation 1 within the ICA framework, demonstrating

that the permutation ambiguity is inherent to Blind Source Separation. This ambiguity is to be expected  $\hat{U}$  in separating the sources we do not seek to impose any restrictions on the order of the separated signals. Thus all permutations of the sources are equally valid.

## 3.5 Preprocessing

Before examining specific ICA algorithms, it is instructive to discuss preprocessing steps that are generally carried out before ICA.

### 3.5.1 Centering

A simple preprocessing step that is commonly performed is to “center” the observation vector  $x$  by subtracting its mean vector  $m = E\{x\}$ . That is then we obtain the centered observation vector,  $x_c$ , as follows:

$$x_c = x - m \quad (16)$$

This step simplifies ICA algorithms by allowing us to assume a zero mean. Once the unmixing matrix has been estimated using the centered data, we can obtain the actual estimates of the independent components as follows:

$$\hat{s}(t) = A^{-1}(x_c + m) \quad (17)$$

From this point on, all observation vectors will be assumed centered. The mixing matrix, on the other hand, remains the same after this preprocessing, so we can always do this without affecting the estimation of the mixing matrix.

### 3.5.2 Whitening

Another step which is very useful in practice is to prewhiten the observation vector  $x$ . Whitening involves linearly transforming the observation vector such that its components are uncorrelated and have unit variance [27]. Let  $x_w$  denote the whitened vector, then it satisfies the following equation:

$$E\{x_w x_w^T\} = I \quad (18)$$

where  $E\{x_w x_w^T\}$  is the covariance matrix of  $x_w$ . Also, since the ICA framework is insensitive to the variances of the independent components, we can assume without loss of generality that the source vector,  $s$ , is white, i.e.  $E\{ss^T\} = I$

A simple method to perform the whitening transformation is to use the eigenvalue decomposition (EVD) [27] of  $x$ . That is, we decompose the covariance matrix of  $x$  as follows:

$$E\{xx^T\} = VDVT \quad (19)$$

where  $V$  is the matrix of eigenvectors of  $E\{xx^T\}$ , and  $D$  is the diagonal matrix of eigenvalues, i.e.  $D =$

$diag\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ . The observation vector can be whitened by the following transformation:

$$x_w = VD^{-1/2}V^T x \tag{20}$$

where the matrix  $D^{-1/2}$  is obtained by a simple component wise operation as  $D^{-1/2} = diag\{\lambda_1^{-1/2}, \lambda_2^{-1/2}, \dots, \lambda_n^{-1/2}\}$ . Whitening transforms the mixing matrix into a new one, which is orthogonal

$$x_w = VD^{-1/2}V^T A s = A_w s \tag{21}$$

hence,

$$\begin{aligned} E\{x_w x_w^T\} &= A_w E\{s s^T\} A_w^T \\ &= A_w A_w^T \\ &= I \end{aligned} \tag{22}$$

Whitening thus reduces the number of parameters to be estimated. Instead of having to estimate the  $n^2$  elements of the original matrix  $A$ , we only need to estimate the new orthogonal mixing matrix, where an orthogonal matrix has  $n(n-1)/2$  degrees of freedom. One can say that whitening solves half of the ICA problem. This is a very useful step as whitening is a simple and efficient process that significantly reduces the computational complexity of ICA. An illustration of the whitening process with simple ICA source separation process is explained in the later section.

### 3.6 Simple Illustrations of ICA

To clarify the concepts discussed in the preceding sections two simple illustrations of ICA are presented here. The results presented below were obtained using the FastICA algorithm, but could equally well have been obtained from any of the numerous ICA algorithms that have been published in the literature (including the Bell and Sejnowski algorithm).

#### 3.6.1 Separation of Two Signals

This section explains the simple ICA source separation process. In this illustration two independent signals,  $s_1$  and  $s_2$ ,

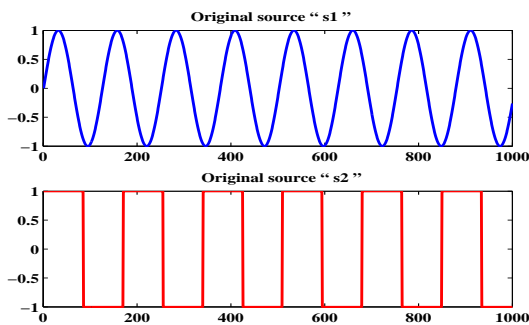


Figure 2: Independent sources  $s_1$  and  $s_2$

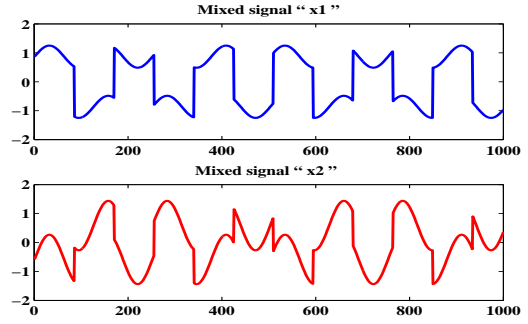


Figure 3: Observed signals,  $x_1$  and  $x_2$ , from an unknown linear mixture of unknown independent components

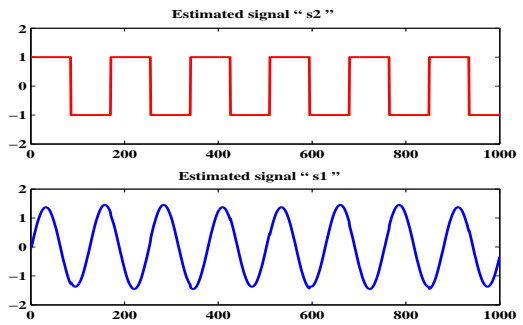


Figure 4: Estimates of independent components

are generated. These signals are shown in Figure 2. The independent components are then mixed according to equation 1 using an arbitrarily chosen mixing matrix  $A$ , where

$$A = \begin{pmatrix} 0.3816 & 0.8678 \\ 0.8534 & -0.5853 \end{pmatrix}$$

The resulting signals from this mixing are shown in Figure 3. Finally, the mixtures  $x_1$  and  $x_2$  are separated using ICA to obtain  $s_1$  and  $s_2$ , shown in Figure 4. By comparing Figure 4 to Figure 2 it is clear that the independent components have been estimated accurately and that the independent components have been estimated without any knowledge of the components themselves or the mixing process.

This example also provides a clear illustration of the scaling and permutation ambiguities discussed in Section 3.4. The amplitudes of the corresponding waveforms in Figures 2 and 4 are different. Thus the estimates of the independent components are some multiple of the independent components of Figure 3, and in the case of  $s_1$ , the scaling factor is negative. The permutation ambiguity is also demonstrated as the order of the independent components has been reversed between Figure 2 and Figure 4.

#### 3.6.2 Illustration of Statistical Independence in ICA

The previous example was a simple illustration of how ICA is used; we start with mixtures of signals and use ICA to

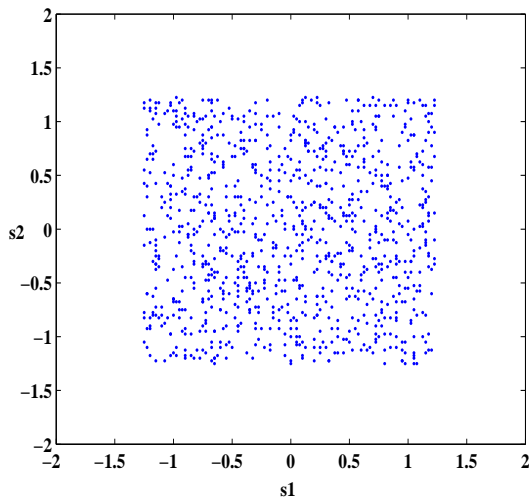


Figure 5: Original sources

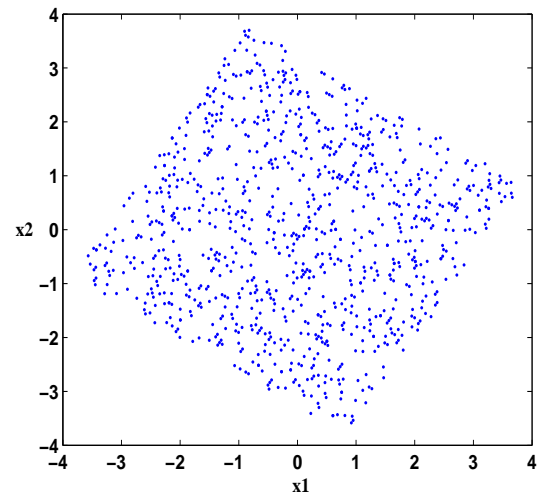


Figure 7: Joint density of whitened signals obtained from whitening the mixed sources

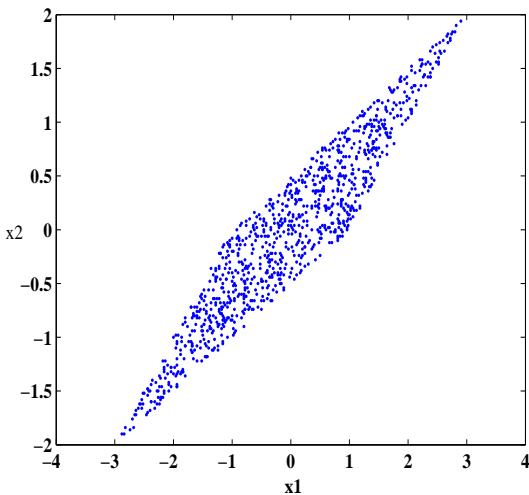


Figure 6: Mixed sources

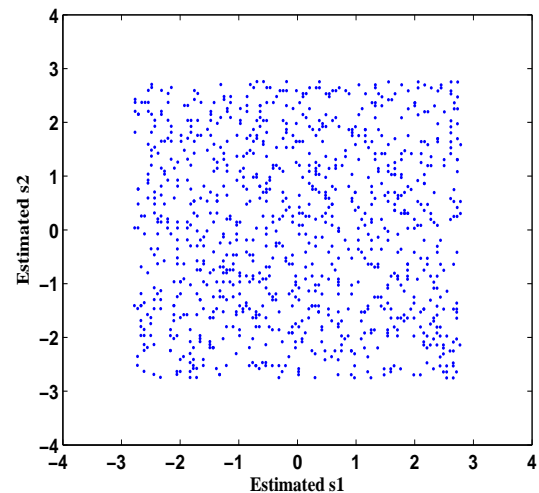


Figure 8: ICA solution (Estimated sources)

separate them. However, this gives no insight into the mechanics of ICA and the close link with statistical independence. We assume that the independent components can be modeled as realizations of some underlying statistical distribution at each time instant (e.g. a speech signal can be accurately modeled as having a Laplacian distribution). One way of visualizing ICA is that it estimates the optimal linear transform to maximise the independence of the joint distribution of the signals  $X_i$ .

The statistical basis of ICA is illustrated more clearly in this example. Consider two random signals which are mixed using the following mixing process:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$$

Figure 5 shows the scatter-plot for original sources  $s_1$  and  $s_2$ . Figure 6 shows the scatter-plot of the mixtures. The

distribution along the axis  $x_1$  and  $x_2$  are now dependent and the form of the density is stretched according to the mixing matrix. From the Figure 6 it is clear that the two signals are not statistically independent because, for example, if  $x_1 = -3$  or  $3$  then  $x_2$  is totally determined. Whitening is an intermediate step before ICA is applied. The joint distribution that results from whitening the signals of Figure 6 is shown in Figure 7. By applying ICA, we seek to transform the data such that we obtain two independent components.

The joint distribution resulting from applying ICA to  $x_1$  and  $x_2$  is shown in Figure 8. This is clearly the joint distribution of two independent, uniformly distributed random variables. Independence can be intuitively confirmed as each random variable is unconstrained regardless of the value of the other random variable (this is not the case for  $x_1$  and  $x_2$ ). The uniformly distributed random variables in

Figure 8 take values between 3 and -3, but due to the scaling ambiguity, we do not know the range of the original independent components. By comparing the whitened data of Figure 7 with Figure 8, we can see that, in this case, pre-whitening reduces ICA to finding an appropriate rotation to yield independence. This is a simplification as a rotation is an orthogonal transformation which requires only one parameter.

The two examples in this section are simple but they illustrate both how ICA is used and the statistical underpinnings of the process. The power of ICA is that an identical approach can be used to address problems of much greater complexity.

### 3.7 ICA Algorithms

There are several ICA algorithms available in literature. However the following three algorithms are widely used in numerous signal processing applications. These include FastICA, JADE, and Infomax. Each algorithm used a different approach to solve equation.

#### 3.7.1 FastICA

FastICA is a fixed point ICA algorithm that employs higher order statistics for the recovery of independent sources. FastICA can estimate ICs one by one (deflation approach) or simultaneously (symmetric approach). FastICA uses simple estimates of Negentropy based on the maximum entropy principle, which requires the use of appropriate nonlinearities for the learning rule of the neural network.

Fixed point algorithm is based on the mutual information. Which can be written as:

$$I(s) = \int f_s(s) \log \frac{f_s(s)}{\prod f_{s_i}(s_i)} ds \quad (23)$$

This measure is kind of distance of independence. Minimising mutual information leads to ICA solution. For the fast ICA algorithm the above equation is re written as

$$I(s) = J(s) - \sum_i J s_i + \frac{1}{2} \log \frac{\prod C_{ii}}{\det C_{ss}} \quad (24)$$

where  $\hat{s} = Wx$ ,  $C_{ss}$  is the correlation matrix, and  $c_{ii}$  is the  $i$ th diagonal element of the correlation matrix. The last term is zero because  $s_i$  are supposed to be uncorrelated. The first term is constant for a problem, because of the invariance in Negentropy. The problem is now reduced to separately maximising the Negentropy of each component.

Estimation of Negentropy is a delicate problem. The papers [28][1] and [2][29]

have addressed this problem. For the general version of fixed point algorithm, the approximation was based on a maximum entropy principle. The algorithm works with whitened data, although aversion of non-whitened data exists.

– Criteria

The maximisation is preferred over the following index

$$J_G(w) = [E\{G(w^T v)\}] - E\{G(v)\}^2 \quad (25)$$

to find one independent component, with  $v$  standard gaussian variable, and  $G$ , the one unit contrast function.

– Update rule

Update rule for the generic algorithm is

$$\begin{aligned} w^* &= E\{vg(w^T v)\} - E\{g'(w^T v)\}w \\ w &= w^* / \|w^*\| \end{aligned} \quad (26)$$

to extract one component. There is symmetric version of the FP algorithm, whose update rule is

$$\begin{aligned} W^* &= E\{g(Wv)v^T\} - \text{Diag}(E\{g'(Wv)\})W \\ W &= (W^*W^{*T})^{-1/2}W^* \end{aligned} \quad (27)$$

where  $\text{Diag}(v)$  is a diagonal matrix with  $\text{Diag}_{ii}(v) = v_i$ .

– Parameters

FastICA uses the following nonlinear parameters for convergence.

$$g(y) = \begin{cases} y^3 \\ \tanh(y) \end{cases} \quad (28)$$

The choice is free except that the symmetric algorithm with  $\tanh$  non linearity does not separate super Gaussian signals. Otherwise the choice can be devoted to the other criteria, for instance the cubic non linearity is faster, whereas the  $\tanh$  linearity is more stable. These questions are addressed in [25]

In practice, the expectations in FastICA must be replaced by their estimates. The natural estimates are of course the corresponding sample means. Ideally, all the data available should be used, but this is often not a good idea because the computations may become too demanding. Then the averages can be estimated using a smaller sample, whose size may have a considerable effect on the accuracy of the final estimates. The sample points should be chosen separately at every iteration. If the convergence is not satisfactory, one may then increase the sample size. This thesis uses FastICA algorithm for all applications.

#### 3.7.2 Infomax

The BSS algorithm, proposed by Bell and Sejnowski, [3], is also a gradient based neural network algorithm, with a learning rule for information maximization of information. Infomax uses higher order statistics for the information maximization. In perfect cases, it does provide the best estimate to ICA components. The strength of this algorithm comes from its direct relationship to information theory.



The algorithm is derived through an information maximisation principle, applied here between the inputs and the non linear outputs. Given the form of joint entropy

$$H(s_1, s_2) = H(s_1) + H(s_2) - I(s_1, s_2) \quad (29)$$

Here for two variables  $s = g(Bx)$ , it is clear that maximising the joint entropy of the outputs amounts to minimising mutual information  $I(y_1, y_2)$ , unless it is more interesting to maximise the individual entropies than to reduce the mutual information. This is the point, where the nonlinear function plays an important role.

The basic idea of the information maximisation is to match the slope of the nonlinear function with the input probability density function. That is

$$s = g(x, \theta) \simeq \int_{-\text{inf}}^x f_x(t) dt \quad (30)$$

In case of perfect matching  $f_s(s)$  looks like an uniform variable, whose entropy is large. If this is not possible because the shapes are different, the best solution found in some case is to mix the input distributions so that the resulting mix matches the slope of the transfer function better than a single input distribution. In this case the algorithm does not converge, and the separation is not achieved.

– Criteria

The algorithm is a stochastic gradient ascent that maximises the joint entropy (Eqn. 12).

– Update rule

In its original form, the update rule is

$$\begin{aligned} \Delta B &= \lambda [[B^T]^{-1} + (1 - 2g(Bx + b_0))x^T] \\ \Delta b &= \lambda [1 - 2g(Bx + b_0)] \end{aligned} \quad (31)$$

– Parameters

The nonlinear function used in the original algorithm is

$$g(s) = \frac{1}{1 + e^{-s}} \quad (32)$$

and in the extended version, it is

$$g(s) = s \pm \tanh(s) \quad (33)$$

where the sign is that of the estimated kurtosis of the signal.

The information maximization algorithm (often referred as infomax) is widely used to separate super-Gaussian sources. Infomax is a gradient-based neural network algorithm, with a learning rule for information maximization. Infomax uses higher order statistics for the information maximization. The information maximization is attained by maximizing the joint entropy of a transformed vector.  $z = g(Wx)$ , where  $g$  is a point wise sigmoidal nonlinear function.

## 4 ICA for different conditions

One of the important conditions of ICA is that the number of sensors should be equal to the number of sources. Unfortunately, the real source separation problem does not always satisfy this constraint. This section focusses on ICA source separation problem under different conditions where the number of sources are not equal to the number of recordings.

### 4.1 Overcomplete ICA

Overcomplete ICA is one of the ICA source separation problem where the number of sources are greater than the number of sensors, i.e ( $n > m$ ). The ideas used for overcomplete ICA originally stem from coding theory, where the task is to find a representation of some signals in a given set of generators which often are more numerous than the signals, hence the term overcomplete basis. Sometimes this representation is advantageous as it uses as few ‘basis’ elements as possible, referred to as sparse coding. Olshausen and Field [30] first put these ideas into an information theoretic context by decomposing natural images into an overcomplete basis. Later, Harpur and Prager [31] and, independently, Olshausen [32] presented a connection between sparse coding and ICA in the square case. Lewicki and Sejnowski [22] then were the first to apply these terms to overcomplete ICA, which was further studied and applied by Lee et al. [33]. De Lathauwer et al. [34] provided an interesting algebraic approach to overcomplete ICA of three sources and two mixtures by solving a system of linear equations in the third and fourth-order cumulants, and Bofill and Zibulevsky [35] treated a special case (‘delta-like’ source distributions) of source signals after Fourier transformation. Overcomplete ICA has major applications in bio signal processing, due to the limited number of electrodes (recordings) compared to the number active muscles (sources) involved (in certain cases unlimited).

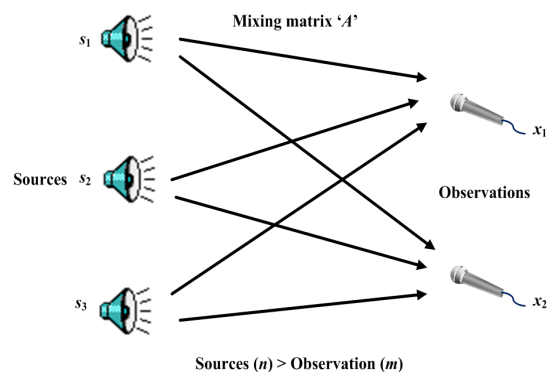


Figure 9: Illustration of “overcomplete ICA”

In overcomplete ICA, the number of sources exceed number of recordings. To analyse this, consider two recordings  $x_1(t)$  and  $x_2(t)$  from three independent sources  $s_1(t)$ ,  $s_2(t)$  and  $s_3(t)$ . The  $x_i(t)$  are then weighted sums

of the  $s_i(t)$ , where the coefficients depend on the distances between the sources and the sensors (refer Figure 9):

$$\begin{aligned} x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) + a_{13}s_3(t) \\ x_2(t) &= a_{21}s_1(t) + a_{22}s_2(t) + a_{23}s_3(t) \end{aligned} \quad (34)$$

The  $a_{ij}$  are constant coefficients that give the mixing weights. The mixing process of these vectors can be represented in the matrix form as (refer Equation 1):

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix}$$

The unmixing process and estimation of sources can be written as (refer Equation 2):

$$\begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix} = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

In this example matrix  $A$  of size  $2 \times 3$  matrix and unmixing matrix  $W$  is of size  $3 \times 2$ . Hence in overcomplete ICA it always results in pseudoinverse. Hence computation of sources in overcomplete ICA requires some estimation processes.

#### 4.1.1 Overcomplete ICA methods

There are two common approaches of solving the overcomplete problem.

- Single step approach where the mixing matrix and the independent sources are estimated at once in a single algorithm
- Two step algorithm where the mixing matrix and the independent component values are estimated with different algorithms.

Lewicki and Sejnowski [22] proposed the single step approach, which is a natural solution to decomposition by finding the maximum a posteriori representation of the data. The prior distribution on the basis function coefficients removes the redundancy in the representation and leads to representations that are sparse and are nonlinear functions of the data. The probabilistic approach to decomposition also leads to a natural method of denoising. From this model, they derived a simple and robust learning algorithm by maximizing the data likelihood over the basis functions. Another approach in single step was proposed by Shriki et al. [36] using recurrent model, i.e., the estimated independent sources are computed taking into account the influence of other independent sources.

One of the disadvantage of single step approach is that it is complex and computationally expensive. Hence many researchers have proposed the two step method, where the mixing matrix is estimated in the first step and the sources are recovered in the next step. Zibulevsky et al. [35] proposed a sparse overcomplete ICA with delta distributions.

Fabian Theis [37, 38] proposed geometric overcomplete ICA. Recently Waheed et. al [39, 40] demonstrated algebraic overcomplete ICA. In this thesis Zibulevsky's sparse overcomplete ICA is utilised, which is explained in the next section.

#### 4.1.2 Sparse overcomplete ICA

Sparse representation of signals which is modeled by matrix factorisation has been receiving a great deal of interest in recent years. The research community has investigated many linear transforms that make audio, video and image data sparse, such as the Discrete Cosine Transform (DCT), the Fourier transform, the wavelet transform and their derivatives. [41]. Chen et al. [42] discussed sparse representations of signals by using large scale linear programming under given overcomplete basis (e.g., wavelets). Olshausen et al. [43] represented sparse coding of images based on maximum posterior approach but it was Zibulevsky et al. [35] who noticed that in the case of sparse sources, their linear mixtures can be easily separated using very simple "geometric" algorithms. Sparse representations can be used in blind source separation. When the sources are sparse, smaller coefficients are more likely and thus for a given data point  $t$ , if one of the sources is significantly larger, the remaining ones are likely to be close to zero. Thus the density of data in the mixture space, besides decreasing with the distance from the origin shows a clear tendency to cluster along the directions of the basis vectors. Sparsity is good in ICA for two reasons. First the statistical accuracy with which the mixing matrix  $A$  can be estimated is a function of how non-Gaussian the source distributions are. This suggests that the sparser the sources are the less data is needed to estimate  $A$ . Secondly the quality of the source estimates given  $A$ , is also better for sparser sources. A signal is considered sparse when values of most of the samples of the signal do not differ significantly from zero. These are from sources that are minimally active. Zibulevsky et al. [35] have demonstrated that when the signals are sparse, and the sources of these are independent, these may be separated even when the number of sources exceeds the number of recordings. [35]. The overcomplete limitation suffered by normal ICA is no longer a limiting factor for signals that are very sparse. Zibulevsky also demonstrated that when the signals are sparse, it is possible to determine the number of independent sources in a mixture of unknown signal numbers.

- Source estimation

The first step in two step approach is source separation. Here the source separation process is explained by taking sparse signal as an example. A signal is considered to be sparse if its pdf is close to Laplacian or super-Gaussian. In this case, the basic ICA model in Equation 1 is modified to have more robust representation which can be expressed as,

$$x = As + \xi \tag{35}$$

where  $\xi$  represents noise in the recordings. It is assumed that the independent sources  $s$  can be sparsely represented in a proper signal dictionary

$$s_i = \sum_{k=1}^K C_i^k \varphi_k \tag{36}$$

where  $\varphi_k$  are the *atoms* or *elements* of the dictionary. Important examples are wavelet-related dictionaries such as wavelet and wavelet packets [41]. Equation 36 can be expressed in matrix notation as

$$s = C\Phi \tag{37}$$

by substituting Equation 37 into 35 gives

$$x = AC\Phi + \xi \tag{38}$$

The goal is to estimate the mixing matrix  $A$  and the coefficients  $C$  at the same time so that  $C$  is as sparse as possible. and  $X \approx AC\Phi$ , given only the observed data  $x$  and the dictionary  $\Phi$

Using maximum a posteriori approach, the above goal can be expressed as

$$\max_{A,C} P(A, C|x) \propto \max_{A,C} P(x|A, C)P(A)P(C) \tag{39}$$

Taking into account Equation 35 and Gaussian noise, the conditional probability  $P(x|A, C)$  can be expressed as

$$P(x|A, C) \propto \prod_i \exp\left[-\frac{(x_i - (AC\Phi)_i)^2}{2\sigma^2}\right] \tag{40}$$

Since  $C$  is assumed to be sparse, it can be approximated with the following pdf

$$p_i(C_i^k) \propto \exp[-(\beta_i h(C_i^k))] \tag{41}$$

and hence

$$p(C) \propto \prod_{i,k} \exp[-(\beta_i h(C_i^k))] \tag{42}$$

Assuming the pdf of  $P(A)$  to be uniform, Equation 39 can now be simplified as

$$\max_{A,C} P(A, C|x) \propto \max_{A,C} P(x|A, C)P(C) \tag{43}$$

Finally, the optimisation problem can be formed by substituting 40 and 42 into 43, taking the logarithm and inverting the sign

$$\max_{A,C} P(A, C|x) \propto \min_{A,C} \frac{1}{2\sigma^2} \|AC\Phi - x\|_F^2 + \sum_{i,k} (\beta_i h(C_i^k)) \tag{44}$$

There are several measures of sparsity. The simplest measure is the  $l_0$  norm. One of the drawback of this measure is that, it is discontinuous and difficult to optimise, and also very sensitive to noise. The closest approximation of  $l_0$  is  $l_1$  norm. The validity of this measure can be shown by simplifying equation 44 under zero noise assumption and under Laplacian prior distributions with  $h(C_i^k) = |C_i^k|$ . Under these assumptions the optimisation problem can be decomposed into  $K$  smaller problems for each data point  $c^k$  at time point  $k = 1 \dots K$  as

$$\min_{c^k} \sum_i |c_i^k| \tag{45}$$

subject to  $Ac^k \varphi_k = x_k$ . If small signal  $s$  is sparse in time domain then  $c^k$  in equation 45 can be updated with  $s^k$ .

$$\min_{s^k} \sum_i |s_i^k| \tag{46}$$

subject to  $As^k = x^k$ . Equation 46 can be formulated as linear programming in basic form as

$$\min_{s^k} c^T s^k \tag{47}$$

subject to  $As^k = x^k, s^k \geq 0$  where  $s^k \Leftrightarrow [u^k; v^k], A \Leftrightarrow [A; -A]$  and  $c \Leftrightarrow [1; 1]$ .

– Estimating the mixing matrix

The second step in two step approach is estimating the mixing matrix. There exists various methods to compute the mixing matrix in sparse overcomplete ICA. The most widely used techniques are:

- (i) C-means clustering
- (ii) Algebraic method and
- (iii) Potential function based method

All the above mentioned methods are based on the clustering principle. The difference is the way they estimate the direction of the clusters. The sparsity of the signal plays an important role for estimating the mixing matrix. A simple illustration that is useful to understand this concept can be found in

## 4.2 Undercomplete ICA

The mixture of unknown sources is referred to as undercomplete when the numbers of recordings  $m$ , more than the number of sources  $n$ . In some applications, it is desired to have more recordings than sources to achieve better separation performance. It is generally believed that with more recordings than the sources, it is always possible to get better estimate of the sources. This is not correct unless prior to separation using ICA, dimensional reduction is conducted. This can be achieved by choosing the same

number of principal recordings as the number of sources discarding the rest. To analyse this, consider three recordings  $x_1(t)$ ,  $x_2(t)$  and  $x_3(t)$  from two independent sources  $s_1(t)$  and  $s_2(t)$ . The  $x_i(t)$  are then weighted sums of the  $s_i(t)$ , where the coefficients depend on the distances between the sources and the sensors (refer Figure 10):

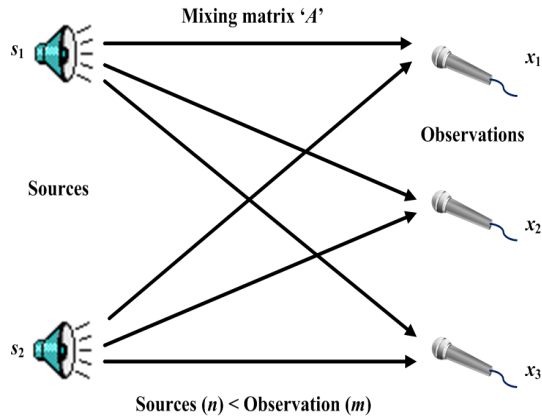


Figure 10: Illustration of “undercomplete ICA”

$$\begin{aligned} x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) \\ x_2(t) &= a_{21}s_1(t) + a_{22}s_2(t) \\ x_3(t) &= a_{31}s_1(t) + a_{32}s_2(t) \end{aligned} \quad (48)$$

The  $a_{ij}$  are constant coefficients that gives the mixing weights. The mixing process of these vectors can be represented in the matrix form as:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$$

The unmixing process using the standard ICA requires a dimensional reduction approach so that, if one of the recordings is reduced then the square mixing matrix is obtained, which can use any standard ICA for the source estimation. For instance one of the recordings say  $x_3$  is redundant then the above mixing process can be written as:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$$

Hence unmixing process can use any standard ICA algorithm using the following:

$$\begin{pmatrix} s_1 \\ s_2 \end{pmatrix} = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

The above process illustrates that, prior to source signal separation using undercomplete ICA, it is important to reduce the dimensionality of the mixing matrix and identify the required and discard the redundant recordings. Principal Component Analysis (PCA) is one of the powerful dimensional reduction method used in signal processing applications, which is explained next.

#### 4.2.1 Undercomplete ICA using dimensional reduction method

When the number of recordings  $n$  are more than the number of sources  $m$ , there must be information redundancy in the recordings. Hence the first step is to reduce the dimensionality of the recorded data. If the dimensionality of the recorded data is equal to that of the sources, then standard ICA methods can be applied to estimate the independent sources. An example of this stages methods is illustrated in [44].

One of the popular method used in dimensional reduction method is PCA. PCA uses the decorrelated method to reduce the recorded data  $x$  using a matrix  $V$

$$z = Vx \quad (49)$$

such that  $Ezz^T = I$ . The transformation matrix  $V$  is given by

$$V = D^{\frac{1}{2}}E^T \quad (50)$$

where  $D$  and  $E$  are the Eigenvalue and Eigenvector decomposition of covariance matrix  $C_x$

$$C_x = ED^{\frac{1}{2}}E^T \quad (51)$$

Now it can be proven that

$$\begin{aligned} E\{zz^T\} &= VE\{xx^T\}V^T \\ &= D^{-1/2}E^T EDE^T ED^{-1/2} \\ &= I \end{aligned} \quad (52)$$

The second stage is using any of the standard ICA algorithms discussed in Section 3.2 to estimate the sources. In fact, whitening process through PCA is standard preprocessing in ICA. It means that applying any standard ICA algorithms that incorporates PCA will automatically reduce the dimension before running ICA.

#### 4.3 Sub band decomposition ICA

Despite the success of using standard ICA in many applications, the basic assumptions of ICA may not hold for certain situations where there may be dependency among the signal sources. The standard ICA algorithms are not able to estimate statistically dependent original sources. One proposed technique [13] is that while there may be a degree of dependency among the wide band source signals, narrow band filtering of these signals can provide independence among these signal sources. This assumption is true when each unknown source can be modeled or represented as a linear combination of narrow-band sub-signals. Sub band decomposition ICA, an extension of ICA, assumes that each source is represented as the sum of some independent subcomponents and dependent subcomponents, which have different frequency bands.

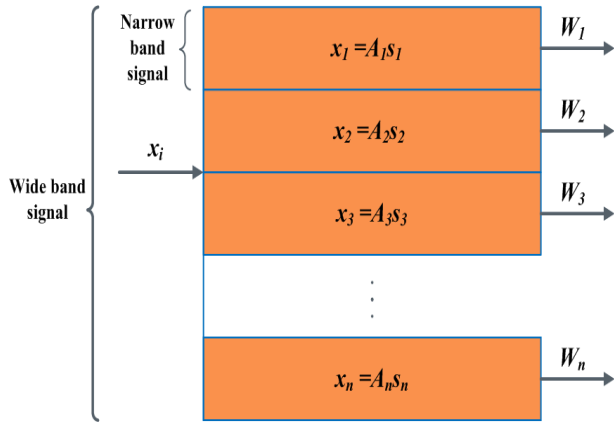


Figure 11: Sub band ICA block diagram.

Such wide-band source signals are a linear decomposition of several narrow-band sub components (refer Figure 11):

$$s(t) = s_1(t) + s_2(t) + s_3(t), \dots, s_n(t) \tag{53}$$

Such decomposition can be modeled in the time, frequency or time frequency domains using any suitable linear transform. A set of unmixing or separating matrices:  $W_1, W_2, W_3, \dots, W_n$  are obtained where  $W_1$  is the unmixing matrix for sensor data  $x_1(t)$  and  $W_n$  is the unmixing matrix for sensor data  $x_n(t)$ . If the specific sub-components of interest are mutually independent for at least two sub-bands, or more generally two subsets of multi-band, say for the sub band “p” and sub band “q” then the global matrix

$$G_{pq} = W_p \times W_q^{-1} \tag{54}$$

will be a sparse generalized permutation matrix  $P$  with special structure with only one non-zero (or strongly dominating) element in each row and each column [27]. This follows from the simple mathematical observation that in such case both matrices  $W_p$  and  $W_q$  represent pseudo-inverses (or true inverse in the case of square matrix) of the same true mixing matrix  $A$  (ignoring non-essential and unavoidable arbitrary scaling and permutation of the columns) and by making an assumption that sources for two multi-frequency sub-bands are independent. This provides the basis for separation of dependent sources using narrow band pass filtered sub band signals for ICA.

### 4.4 Multi run ICA

One of the most effective ways of modeling vector data for unsupervised pattern classification or coding is to assume that the observations are the result of randomly picking out of a fixed set of different distributions. ICA is an iterative BSS technique. At each instance original signals are estimated from the mixed data. The quality of estimation of the original signals depends mainly on the unmixing matrix  $W$ . Due to the randomness associated with the estimation

of the unmixing matrix and the iterative process, there is a randomness associated with the quality of separation.

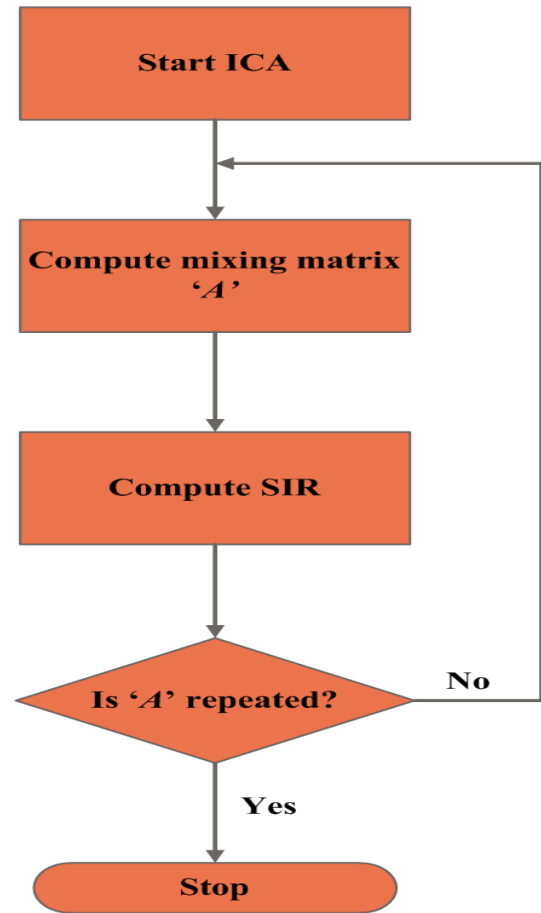


Figure 12: Multi run ICA mixing matrix computation flow chart

Multi run ICA has been proposed to overcome this associated randomness. [45]. It is the process where the ICA algorithm will be computed many times; at each instance different mixing matrices will be estimated.  $A_1, A_2, \dots, A_n$ . Since it is an iterative technique with inbuilt quantisation, repeat analysis yields similarity matrices at some stage. Hence mixing matrices  $A_1, A_2$  etc, will repeat after certain iterations. To estimate the sources from the mixed data ICA requires just one mixing matrix, the best unmixing matrix would give clear source separation, hence the selection of the best matrix is the key criterion in multi run ICA. There exists several methods to compute the quality of the mixing matrices, they are

- Signal to Noise Ratio (SNR)
- Signal to Interference Ratio (SIR)
- Signal to Distortion Ratio (SDR) and
- Signal to Artefacts Ratio (SAR)

In bio signal and audio applications, SIR has found to be a popular tool to measure the quality separation. Once the

best unmixing matrix is estimated, then any normal ICA method can be used for source separation. The multi run ICA computational process flow chart is shown in Figure 12.

## 5 Applications of ICA

The success of ICA in source separation has resulted in a number of practical applications. These includes,

- Machine fault detection [46, 47, 48, 49]
- Seismic monitoring [50, 51]
- Reflection canceling [52, 53]
- Finding hidden factors in financial data [54, 55, 56]
- Text document analysis [4, 5, 6]
- Radio communications [57, 58]
- Audio signal processing [20, 13]
- Image processing [13, 14, 59, 60, 61, 62, 63]
- Data mining [64]
- Time series forecasting [65]
- Defect detection in patterned display surfaces [66, ?]
- Bio medical signal processing [7, 67, 8, 9, 10, 11, 12, 68, 69].

Some of the major applications are explained in detail next:

### 5.1 Biomedical Applications of ICA

Exemplary ICA applications in biomedical problems include the following:

- Fetal Electrocardiogram extraction, i.e removing/filtering maternal electrocardiogram signals and noise from fetal electrocardiogram signals [70, 71].
- Enhancement of low level Electrocardiogram components [70, 71]
- Separation of transplanted heart signals from residual original heart signals [72]
- Separation of low level myoelectric muscle activities to identify various gestures [73, 74, 75, 76]

One successful and promising application domain of blind signal processing includes those biomedical signals acquired using multi-electrode devices: Electrocardiography (ECG), [77, 70, 72, 71, 78, 79, 69], Electroencephalography (EEG)[70, 71, 72, 80, 81, 82], Magnetoencephalography (MEG) [83, 84, 85, 86, 80, 87] and sEMG. Surface EMG is an indicator of muscle activity and related to body movement and posture. It has major applications in biosignal processing, next section explains sEMG and its applications.

### 5.2 Telecommunications

Telecommunication is one of the emerging application with respect to ICA, it has major application in code Division Multiple Access (CDMA) mobile communications. This problem is semi-blind, in the sense that certain additional prior information is available on the CDMA data model [88]. But the number of parameters to be estimated is often so high that suitable BSS, techniques taking into account the available prior knowledge, provide a clear performance improvement over more traditional estimation techniques.

### 5.3 Feature extraction

ICA is successfully applied for face recognition and lip reading. The goal in the face recognition is to train a system that can recognise 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. Barlett and Sejnowski compare the face recognition performance of PCA and ICA for two different tasks:

1. different pose and
2. different lighting conditions

they show that for both the tasks, ICA outperforms PCA.

### 5.4 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 complimentary 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 [89] have applied the ICA for improving the fusion of multimodal surveillance images in sensor networks. ICA is also used for robust speech recognition using various sensor combinations

### 5.5 Audio signal processing

One of the most practical uses for BSS is in the audio world. It has been used for noise removal without the need of filters or Fourier transforms, which leads to simpler processing methods. There are various problems associated with noise removal in this way, but these can most likely be attributed to the relative infancy of the BSS field and such limitations will be reduced as research increases in this field [90, 25].

Audio source separation is the problem of automated separation of audio sources present in a room, using a set of differently placed microphones, capturing the auditory

scene. The whole problem resembles the task a human listener can solve in a cocktail party situation, where using two sensors (ears), the brain can focus on a specific source of interest, suppressing all other sources present (also known as cocktail party problem) [20, 25].

## 5.6 Image Processing

Recently, Independent Component Analysis (ICA) has been proposed as a generic statistical model for images [90, 59, 60, 61, 62, 63]. It is aimed at capturing the statistical structure in images that is beyond second order information, by exploiting higher-order statistical structure in data. ICA finds a linear non orthogonal coordinate system in multivariate data determined by second- and higher-order statistics. The goal of ICA is to linearly transform the data such that the transformed variables are as statistically independent from each other as possible. ICA generalizes PCA and, like PCA, has proven a useful tool for finding structure in data. Bell and Sejnowski proposed a method to extract features from natural scenes by assuming linear image synthesis model [90]. In their model, a set of digitized natural images were used. they considered each patch of an image as a linear combination of several underlying basic functions. Later Lee et al [91] proposed an image processing algorithm, which estimates the data density in each class by using parametric nonlinear functions that fit to the non-Gaussian structure of the data. They showed a significant improvement in classification accuracy over standard Gaussian mixture models. Recently Antoniol et al [92] demonstrated that the ICA model can be a suitable tool for learning a vector base for feature extraction to design a feature based data dependent approach that can be efficiently adopted for image change detection. In addition ICA features are localized and oriented and sensitive to lines and edges of varying thickness of images. Furthermore the sparsity of ICA coefficients should be pointed out. It is expected that suitable soft-thresholding on the ICA coefficients leads to efficient reduction of Gaussian noise [60, 62, 63].

## 6 Conclusions

This paper has introduced the fundamentals of BSS and ICA. The mathematical framework of the source mixing problem that BSS/ICA addresses was examined in some detail, as was the general approach to solving BSS/ICA. As part of this discussion, some inherent ambiguities of the BSS/ICA framework were examined as well as the two important preprocessing steps of centering and whitening. Specific details of the approach to solving the mixing problem were presented and two important ICA algorithms were discussed in detail. Finally, the application domains of this novel technique are presented. Some of the futuristic works on ICA techniques, which need further investigation are discussed. The material covered in this paper is important not only to understand the algorithms used to perform

BSS/ICA, but it also provides the necessary background to understand extensions to the framework of ICA for future researchers.

## References

- [1] L. Tong, Liu, V. C. Soon, and Y. F. Huang, "Indeterminacy and identifiability of blind identification," *Circuits and Systems, IEEE Transactions on*, vol. 38, no. 5, pp. 499–509, 1991.
- [2] C. Jutten and J. Karhunen, "Advances in blind source separation (bss) and independent component analysis (ica) for nonlinear mixtures." *Int J Neural Syst*, vol. 14, no. 5, pp. 267–292, October 2004.
- [3] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution." *Neural Comput*, vol. 7, no. 6, pp. 1129–1159, November 1995.
- [4] Kolenda, *Independent components in text*, ser. Advances in Independent Component Analysis. Springer-Verlag, 2000, pp. 229–250.
- [5] E. Bingham, J. Kuusisto, and K. Lagus, "Ica and som in text document analysis," in *SIGIR '02: Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2002, pp. 361–362.
- [6] Q. Pu and G.-W. Yang, "Short-text classification based on ica and lsa," *Advances in Neural Networks - ISNN 2006*, pp. 265–270, 2006.
- [7] C. J. James and C. W. Hesse, "Independent component analysis for biomedical signals," *Physiological Measurement*, vol. 26, no. 1, pp. R15+, 2005.
- [8] B. Azzerboni, M. Carpentieri, F. La Foresta, and F. C. Morabito, "Neural-ica and wavelet transform for artifacts removal in surface emg," in *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, vol. 4, 2004, pp. 3223–3228 vol.4.
- [9] F. De Martino, F. Gentile, F. Esposito, M. Balsi, F. Di Salle, R. Goebel, and E. Formisano, "Classification of fmri independent components using ic-fingerprints and support vector machine classifiers," *NeuroImage*, vol. 34, pp. 177–194, 2007.
- [10] T. Kumagai and A. Utsugi, "Removal of artifacts and fluctuations from meg data by clustering methods," *Neurocomputing*, vol. 62, pp. 153–160, December 2004.
- [11] Y. Zhu, T. L. Chen, W. Zhang, T.-P. Jung, J.-R. Duan, S. Makeig, and C.-K. Cheng, "Noninvasive study of the human heart using independent component analysis," in *BIBE '06: Proceedings of the Sixth IEEE*

- Symposium on BionInformatics and BioEngineering*. IEEE Computer Society, 2006, pp. 340–347.
- [12] J. Enderle, S. M. Blanchard, and J. Bronzino, Eds., *Introduction to Biomedical Engineering, Second Edition*. Academic Press, April 2005.
- [13] A. Cichocki and S.-I. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. John Wiley & Sons, Inc., 2002.
- [14] Q. Zhang, J. Sun, J. Liu, and X. Sun, “A novel ica-based image/video processing method,” 2007, pp. 836–842.
- [15] Hansen, *Blind separation of noisy image mixtures*. Springer-Verlag, 2000, pp. 159–179.
- [16] D. J. C. Mackay, “Maximum likelihood and covariant algorithms for independent component analysis,” University of Cambridge, London, Tech. Rep., 1996.
- [17] Sorenson, “Mean field approaches to independent component analysis,” *Neural Computation*, vol. 14, pp. 889–918, 2002.
- [18] Kabâ’ân, “Clustering of text documents by skewness maximization,” 2000, pp. 435–440.
- [19] T. W. Lee, “Blind separation of delayed and convolved sources,” 1997, pp. 758–764.
- [20] —, *Independent component analysis: theory and applications*. Kluwer Academic Publishers, 1998.
- [21] H. Attias and C. E. Schreiner, “Blind source separation and deconvolution: the dynamic component analysis algorithm,” *Neural Comput.*, vol. 10, no. 6, pp. 1373–1424, August 1998.
- [22] M. S. Lewicki and T. J. Sejnowski, “Learning overcomplete representations,” *Neural Comput*, vol. 12, no. 2, pp. 337–365, February 2000.
- [23] A. Hyvarinen, R. Cristescu, and E. Oja, “A fast algorithm for estimating overcomplete ica bases for image windows,” in *Neural Networks, 1999. IJCNN '99. International Joint Conference on*, vol. 2, 1999, pp. 894–899 vol.2.
- [24] T. W. Lee, M. S. Lewicki, and T. J. Sejnowski, “Unsupervised classification with non-gaussian mixture models using ica,” in *Proceedings of the 1998 conference on Advances in neural information processing systems*. Cambridge, MA, USA: MIT Press, 1999, pp. 508–514.
- [25] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Wiley-Interscience, May 2001.
- [26] J. V. Stone, *Independent Component Analysis : A Tutorial Introduction (Bradford Books)*. The MIT Press, September 2004.
- [27] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*. Cambridge, UK, 2000.
- [28] P. Comon, “Independent component analysis, a new concept?” *Signal Processing*, vol. 36, no. 3, pp. 287–314, april 1994.
- [29] A. Hyvrinen, “New approximations of differential entropy for independent component analysis and projection pursuit,” in *NIPS '97: Proceedings of the 1997 conference on Advances in neural information processing systems 10*. MIT Press, 1998, pp. 273–279.
- [30] Olshausen, “Sparse coding of natural images produces localized, oriented, bandpass receptive fields,” Department of Psychology, Cornell University, Tech. Rep., 1995.
- [31] G. F. Harpur and R. W. Prager, “Development of low entropy coding in a recurrent network,” *Network (Bristol, England)*, vol. 7, no. 2, pp. 277–284, May 1996.
- [32] B. A. Olshausen, “Learning linear, sparse, factorial codes,” Tech. Rep., 1996.
- [33] T. W. Lee, M. Girolami, M. S. Lewicki, and T. J. Sejnowski, “Blind source separation of more sources than mixtures using overcomplete representations,” *Signal Processing Letters, IEEE*, vol. 6, no. 4, pp. 87–90, 2000.
- [34] D. Lathauwer, P. L. Comon, B. De Moor, and J. Vandewalle, “Ica algorithms for 3 sources and 2 sensors,” in *Higher-Order Statistics, 1999. Proceedings of the IEEE Signal Processing Workshop on*, 1999, pp. 116–120.
- [35] Bofill, “Blind separation of more sources than mixtures using sparsity of their short-time fourier transform,” Pajunen, Ed., 2000, pp. 87–92.
- [36] O. Shriki, H. Sompolinsky, and D. D. Lee, “An information maximization approach to overcomplete and recurrent representations,” in *In Advances in Neural Information Processing Systems*, vol. 14, 2002, pp. 612–618.
- [37] F. J. Theis, E. W. Lang, T. Westenhuber, and C. G. Puntonet, “Overcomplete ica with a geometric algorithm,” in *ICANN '02: Proceedings of the International Conference on Artificial Neural Networks*. Springer-Verlag, 2002, pp. 1049–1054.
- [38] F. J. Theis and E. W. Lang, “Geometric overcomplete ica,” in *Proc. of ESANN 2002*, 2002, pp. 217–223.



- [39] K. Waheed and F. M. Salem, “Algebraic independent component analysis: an approach for separation of overcomplete speech mixtures,” in *Neural Networks, 2003. Proceedings of the International Joint Conference on*, vol. 1, 2003, pp. 775–780 vol.1.
- [40] —, “Algebraic independent component analysis,” in *Robotics, Intelligent Systems and Signal Processing, 2003. Proceedings. 2003 IEEE International Conference on*, vol. 1, 2003, pp. 472–477 vol.1.
- [41] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 1998.
- [42] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.
- [43] B. A. Olshausen and D. J. Field, “Sparse coding with an overcomplete basis set: a strategy employed by v1?” *Vision Res.*, vol. 37, no. 23, pp. 3311–3325, December 1997.
- [44] M. Joho, H. Mathis, and R. Lambert, “Overdetermined blind source separation: Using more sensors than source signals in a noisy mixture,” 2000.
- [45] G. R. Naik, D. K. Kumar, and M. Palaniswami, “Multi run ica and surface emg based signal processing system for recognising hand gestures,” in *Computer and Information Technology, 2008. CIT 2008. 8th IEEE International Conference on*, 2008, pp. 700–705.
- [46] A. Ypma, D. M. J. Tax, and R. P. W. Duin, “Robust machine fault detection with independent component analysis and support vector data description,” in *Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop*, 1999, pp. 67–76.
- [47] Z. Li, Y. He, F. Chu, J. Han, and W. Hao, “Fault recognition method for speed-up and speed-down process of rotating machinery based on independent component analysis and factorial hidden markov model,” *Journal of Sound and Vibration*, vol. 291, no. 1-2, pp. 60–71, March 2006.
- [48] M. Kano, S. Tanaka, S. Hasebe, I. Hashimoto, and H. Ohno, “Monitoring independent components for fault detection,” *AIChE Journal*, vol. 49, no. 4, pp. 969–976, 2003.
- [49] L. Zhonghai, Z. Yan, J. Liying, and Q. Xiaoguang, “Application of independent component analysis to the aero-engine fault diagnosis,” in *2009 Chinese Control and Decision Conference*. IEEE, June 2009, pp. 5330–5333.
- [50] de La, C. G. Puntonet, J. M. Górriz, and I. Lloret, “An application of ica to identify vibratory low-level signals generated by termites,” 2004, pp. 1126–1133.
- [51] F. Acernese, A. Ciaramella, S. De Martino, M. Falanga, C. Godano, and R. Tagliaferri, “Polarisation analysis of the independent components of low frequency events at stromboli volcano (eolian islands, italy),” *Journal of Volcanology and Geothermal Research*, vol. 137, no. 1-3, pp. 153–168, September 2004.
- [52] H. Farid and E. H. Adelson, “Separating reflections and lighting using independent components analysis,” *cvpr*, vol. 01, 1999.
- [53] M. Yamazaki, Y.-W. Chen, and G. Xu, “Separating reflections from images using kernel independent component analysis,” in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 3, 2006, pp. 194–197.
- [54] M. Coli, R. Di Nisio, and L. Ippoliti, “Exploratory analysis of financial time series using independent component analysis,” in *Information Technology Interfaces, 2005. 27th International Conference on*, 2005, pp. 169–174.
- [55] E. H. Wu and P. L. Yu, “Independent component analysis for clustering multivariate time series data,” 2005, pp. 474–482.
- [56] S.-M. Cha and L.-W. Chan, “Applying independent component analysis to factor model in finance,” in *IDEAL '00: Proceedings of the Second International Conference on Intelligent Data Engineering and Automated Learning, Data Mining, Financial Engineering, and Intelligent Agents*. Springer-Verlag, 2000, pp. 538–544.
- [57] R. Cristescu, T. Ristaniemi, J. Joutsensalo, and J. Karhunen, “Cdma delay estimation using fast ica algorithm,” vol. 2, 2000, pp. 1117–1120 vol.2.
- [58] J. P. Huang and J. Mar, “Combined ica and fca schemes for a hierarchical network,” *Wirel. Pers. Commun.*, vol. 28, no. 1, pp. 35–58, January 2004.
- [59] O. Déniz, M. Castrillón, and M. Hernández, “Face recognition using independent component analysis and support vector machines,” *Pattern Recogn. Lett.*, vol. 24, no. 13, pp. 2153–2157, 2003.
- [60] S. Fiori, “Overview of independent component analysis technique with an application to synthetic aperture radar (sar) imagery processing,” *Neural Netw.*, vol. 16, no. 3-4, pp. 453–467, 2003.
- [61] H. Wang, Y. Pi, G. Liu, and H. Chen, “Applications of ica for the enhancement and classification of polarimetric sar images,” *Int. J. Remote Sens.*, vol. 29, no. 6, pp. 1649–1663, 2008.
- [62] M. S. Karoui, Y. Deville, S. Hosseini, A. Ouamri, and D. Ducrot, “Improvement of remote sensing multispectral image classification by using independent

- component analysis,” in *2009 First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*. IEEE, August 2009, pp. 1–4.
- [63] L. Xiaochun and C. Jing, “An algorithm of image fusion based on ica and change detection,” in *Proceedings 7th International Conference on Signal Processing, 2004. Proceedings. ICSP '04. 2004*. IEEE, 2004, pp. 1096–1098.
- [64] J.-H. H. Lee, S. Oh, F. A. Jolesz, H. Park, and S.-S. S. Yoo, “Application of independent component analysis for the data mining of simultaneous eeg-fmri: preliminary experience on sleep onset.” *The International journal of neuroscience*, vol. 119, no. 8, pp. 1118–1136, 2009. [Online]. Available: <http://view.ncbi.nlm.nih.gov/pubmed/19922343>
- [65] C.-J. Lu, T.-S. Lee, and C.-C. Chiu, “Financial time series forecasting using independent component analysis and support vector regression,” *Decis. Support Syst.*, vol. 47, no. 2, pp. 115–125, 2009.
- [66] D.-M. Tsai, P.-C. Lin, and C.-J. Lu, “An independent component analysis-based filter design for defect detection in low-contrast surface images,” *Pattern Recogn.*, vol. 39, no. 9, pp. 1679–1694, 2006.
- [67] F. Castells, J. Igual, J. Millet, and J. J. Rieta, “Atrial activity extraction from atrial fibrillation episodes based on maximum likelihood source separation,” *Signal Process.*, vol. 85, no. 3, pp. 523–535, 2005.
- [68] H. Safavi, N. Correa, W. Xiong, A. Roy, T. Adali, V. R. Korostyshevskiy, C. C. Whisnant, and F. Seillier-Moiseiwitsch, “Independent component analysis of 2-d electrophoresis gels,” *ELECTROPHORESIS*, vol. 29, no. 19, pp. 4017–4026, 2008.
- [69] R. Llinares and J. Igual, “Application of constrained independent component analysis algorithms in electrocardiogram arrhythmias,” *Artif. Intell. Med.*, vol. 47, no. 2, pp. 121–133, 2009.
- [70] E. Niedermeyer and F. L. Da Silva, *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams and Wilkins; 4th edition, January 1999.
- [71] J. C. Rajapakse, A. Cichocki, and Sanchez, “Independent component analysis and beyond in brain imaging: Eeg, meg, fmri, and pet,” in *Neural Information Processing, 2002. ICONIP '02. Proceedings of the 9th International Conference on*, vol. 1, 2002, pp. 404–412 vol.1.
- [72] J. Wisbeck, A. Barros, and R. Ojeda, “Application of ica in the separation of breathing artifacts in eeg signals,” 1998.
- [73] S. Calinon and A. Billard, “Recognition and reproduction of gestures using a probabilistic framework combining pca, ica and hmm,” in *ICML '05: Proceedings of the 22nd international conference on Machine learning*. ACM, 2005, pp. 105–112.
- [74] M. Kato, Y.-W. Chen, and G. Xu, “Articulated hand tracking by pca-ica approach,” in *FGR '06: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*. IEEE Computer Society, 2006, pp. 329–334.
- [75] G. R. Naik, D. K. Kumar, V. P. Singh, and M. Palaniswami, “Hand gestures for hci using ica of emg,” in *VisHCI '06: Proceedings of the HCSNet workshop on Use of vision in human-computer interaction*. Australian Computer Society, Inc., 2006, pp. 67–72.
- [76] G. R. Naik, D. K. Kumar, H. Weghorn, and M. Palaniswami, “Subtle hand gesture identification for hci using temporal decorrelation source separation bss of surface emg,” in *Digital Image Computing Techniques and Applications, 9th Biennial Conference of the Australian Pattern Recognition Society on*, 2007, pp. 30–37.
- [77] M. Scherg and D. Von Cramon, “Two bilateral sources of the late aep as identified by a spatio-temporal dipole model.” *Electroencephalogr Clin Neurophysiol*, vol. 62, no. 1, pp. 32–44, January 1985.
- [78] R. Phlypo, V. Zarzoso, P. Comon, Y. D’Asseler, and I. Lemahieu, “Extraction of atrial activity from the eeg by spectrally constrained ica based on kurtosis sign,” in *ICA'07: Proceedings of the 7th international conference on Independent component analysis and signal separation*. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 641–648.
- [79] J. Oster, O. Pietquin, R. Abächerli, M. Kraemer, and J. Felblinger, “Independent component analysis-based artefact reduction: application to the electrocardiogram for improved magnetic resonance imaging triggering,” *Physiological Measurement*, vol. 30, no. 12, pp. 1381–1397, December 2009. [Online]. Available: <http://dx.doi.org/10.1088/0967-3334/30/12/007>
- [80] R. Vigário, J. Särelä, V. Jousmäki, M. Hämäläinen, and E. Oja, “Independent component approach to the analysis of eeg and meg recordings.” *IEEE transactions on bio-medical engineering*, vol. 47, no. 5, pp. 589–593, May 2000.
- [81] J. Onton, M. Westerfield, J. Townsend, and S. Makeig, “Imaging human eeg dynamics using independent component analysis,” *Neuroscience & Biobehavioral Reviews*, vol. 30, no. 6, pp. 808–822, 2006.

- [82] B. Jervis, S. Belal, K. Camilleri, T. Cassar, C. Bigan, D. E. J. Linden, K. Michalopoulos, M. Zervakis, M. Besleaga, S. Fabri, and J. Muscat, "The independent components of auditory p300 and cnv evoked potentials derived from single-trial recordings," *Physiological Measurement*, vol. 28, no. 8, pp. 745–771, August 2007.
- [83] J. C. Mosher, P. S. Lewis, and R. M. Leahy, "Multiple dipole modeling and localization from spatio-temporal meg data," *Biomedical Engineering, IEEE Transactions on*, vol. 39, no. 6, pp. 541–557, 1992.
- [84] M. Hämmäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa, "Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain," *Reviews of Modern Physics*, vol. 65, no. 2, pp. 413+, April 1993.
- [85] A. C. Tang and B. A. Pearlmutter, "Independent components of magnetoencephalography: localization," pp. 129–162, 2003.
- [86] J. Parra, S. N. Kalitzin, and Lopes, "Magnetoencephalography: an investigational tool or a routine clinical technique?" *Epilepsy & Behavior*, vol. 5, no. 3, pp. 277–285, June 2004.
- [87] K. Petersen, L. K. Hansen, T. Kolenda, and E. Rostrup, "On the independent components of functional neuroimages," in *Third International Conference on Independent Component Analysis and Blind Source Separation*, 2000, pp. 615–620.
- [88] T. Ristaniemi and J. Joutsensalo, "the performance of blind source separation in cdma downlink," 1999.
- [89] N. Cvejic, D. Bull, and N. Canagarajah, "Improving fusion of surveillance images in sensor networks using independent component analysis," *Consumer Electronics, IEEE Transactions on*, vol. 53, no. 3, pp. 1029–1035, 2007.
- [90] A. J. Bell and T. J. Sejnowski, "The "independent components" of natural scenes are edge filters." *Vision Res*, vol. 37, no. 23, pp. 3327–3338, December 1997.
- [91] T.-W. W. Lee and M. S. Lewicki, "Unsupervised image classification, segmentation, and enhancement using ica mixture models." *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 11, no. 3, pp. 270–279, 2002.
- [92] G. Antoniol, M. Ceccarelli, P. Petrillo, and A. Petrosino, "An ica approach to unsupervised change detection in multispectral images," in *Biological and Artificial Intelligence Environments*, B. Apolloni, M. Marinaro, and R. Tagliaferri, Eds. Springer Netherlands, 2005, ch. 35, pp. 299–311.

