1

Introduction: Independent Component Analysis

Ganesh R. Naik
RMIT University, Melbourne
Australia

1. Introduction

Consider a situation in which we have a number of sources emitting signals which are interfering with one another. Familiar 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 (BSS). 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 Herault, J. & Jutten, C. (1987) 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 Bell & Sejnowski (1995) published a relatively simple approach to the problem named infomax, that many became aware of the potential of Independent component analysis (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 Azzerboni et al. (2004); Bingham et al. (2002); Cichocki & Amari (2002); De Martino et al. (2007); Enderle et al. (2005); James & Hesse (2005); Kolenda (2000); Kumagai & Utsugi (2004); Pu & Yang (2006); Zhang et al. (2007); Zhu et al. (2006).

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 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, Hansen (2000); Mackay (1996), nontrivial source distributions, Kab'an (2000); Sorenson (2002), convolutive mixtures Attias & Schreiner (1998); Lee (1997; 1998), time dependency, underdetermined sources Hyvarinen et al. (1999); Lewicki & Sejnowski (2000), mixture and classification of independent component Kolenda (2000); Lee et al. (1999). A general introduction and overview can be found in Hyvarinen et al. (2001).

1.1 ICA model

ICA is a statistical technique, perhaps the most widely used, for solving the blind source separation problem Hyvarinen et al. (2001); Stone (2004). In this section, we present the basic Independent Component Analysis model and show under which conditions its parameters can be estimated. 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, $s_i(t), i = 1, ..., N$. 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 $x_i(t), i = 1, ..., N$ 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:

$$x(t) = As(t) \tag{1}$$

where $\boldsymbol{A}$ is an unknown matrix called the *mixing matrix* and $\boldsymbol{x}$(t), $\boldsymbol{s}$(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 Şunmixing matrixŤ $\boldsymbol{W}$, where, $W = A^{-1}$.

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

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

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.

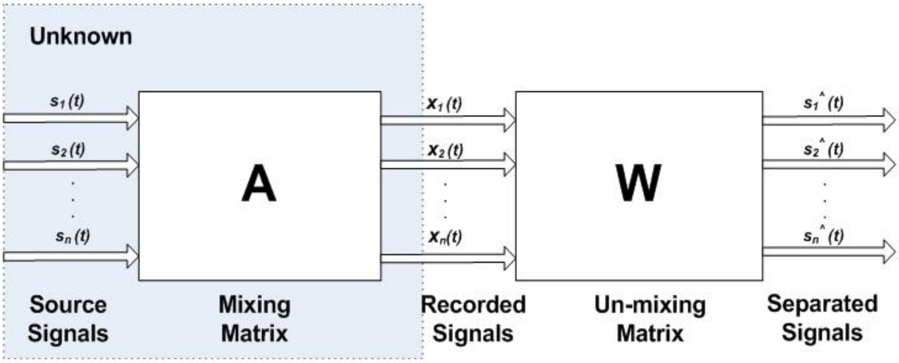

Fig. 1. Blind source separation (BSS) block diagram. $s(t)$ are the sources. $x(t)$ are the recordings, $\hat{s}(t)$ are the estimated sources A is mixing matrix and W is un-mixing matrix

1.2 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.

1.2.1 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{.} is the expectation operator. In the following section we use the above properties to explain the relationship between uncorrelated and independence.

1.2.2 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.

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

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. How ever 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.

1.2.3 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

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

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, $E\{s^4\} = 3(E\{s^4\})^2$ 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

$$kurt(s_1 \pm s_2) = kurt(s_1) \pm kurt(s_2) \tag{8}$$

and

$$kurt(\alpha s_1) = \alpha^4 kurt(s_1) \tag{9}$$

where α 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. It 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

$$H(S) = -\sum P(S = a_i) log P(S = a_i) \tag{10}$$

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

$$H(S) = -\int p(s) log p(s) ds \tag{11}$$

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

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

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.

1.2.4 ICA assumptions

- *The sources being considered are statistically independent*

The first assumption is fundamental to ICA. As discussed in previous section, 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 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. 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.

1.2.5 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

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

where a_j denotes the jth 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.

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

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.

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

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 Ű 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.

1.3 Preprocessing

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

1.3.1 Centering

A simple preprocessing step that is commonly performed is to ŞcenterŤ 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:

$$x_c = x - m \tag{16}$$

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:

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

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.

1.3.2 Whitening

Another step which is very useful in practice is to pre-whiten 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:

$$E\{x_w x_w^T\} = I \tag{18}$$

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:

$$E\{xx^T\} = VDV^T \tag{19}$$

where V is the matrix of eigenvectors of $E\{xx^T\}$, and D is the diagonal matrix of eigenvalues, i.e. $D = diag\{\lambda_1, \lambda_2, ..., \lambda_n\}$. 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 $D^{-1/2} = diag\{\lambda_1^{-1/2}, \lambda_2^{-1/2}, ..., \lambda_n^{-1/2}\}$. Whitening transforms the mixing matrix into a new one, which is orthogonal

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

hence,

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

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 following section.

1.4 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 Sejnowsiki algorithm).

1.4.1 Separation of two signals

This section explains the simple ICA source separation process. In this illustration two independent signals, s_1 and s_2, are generated. These signals are shown in Figure2. The independent components are then mixed according to equation 1 using an arbitrarily chosen mixing matrix A, where

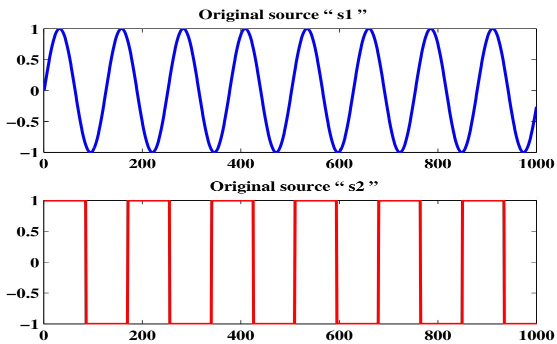

Fig. 2. Independent sources $s1$ and $s2$

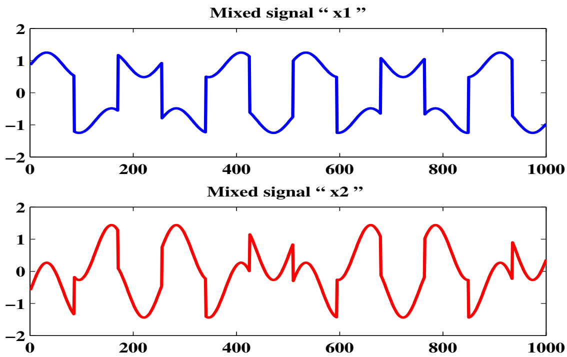

Fig. 3. Observed signals, $x1$ and $x2$, from an unknown linear mixture of unknown independent components

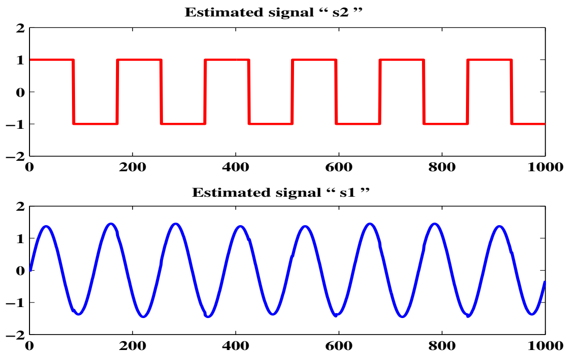

Fig. 4. Estimates of independent components

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

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 previously. 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 $s1$, 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.

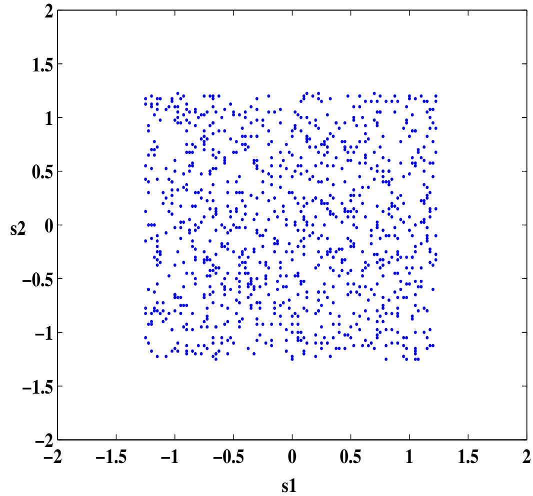

Fig. 5. Original sources

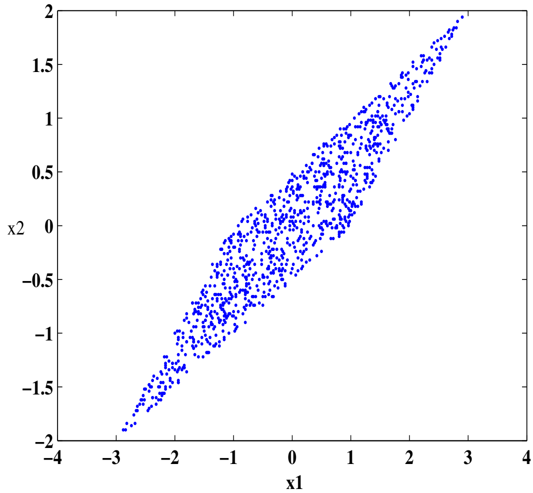

Fig. 6. Mixed sources

1.4.2 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 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

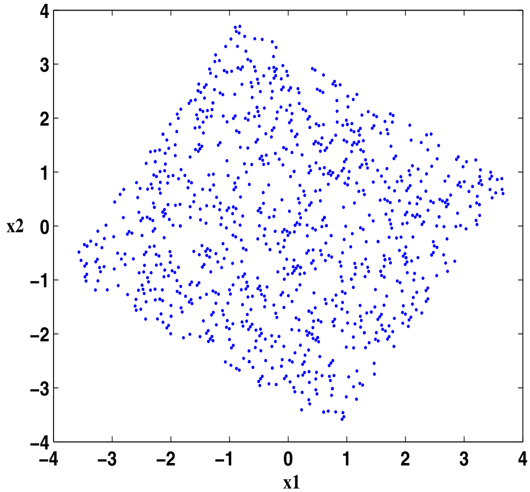

Fig. 7. Joint density of whitened signals obtained from whitening the mixed sources

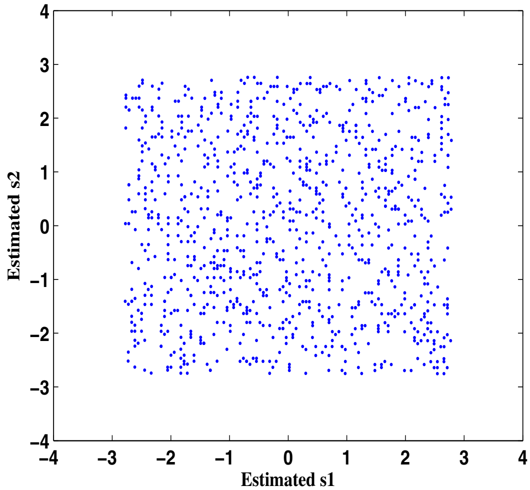

Fig. 8. ICA solution (Estimated sources)

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 7. 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.

2. 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.

2.1 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 Olshausen (1995) first put these ideas into an information theoretic context by decomposing natural images into an overcomplete basis. Later, Harpur and Prager Harpur & Prager (1996) and, independently, Olshausen Olshausen (1996) presented a connection between sparse coding and ICA in the square case. Lewicki and Sejnowski Lewicki & Sejnowski (2000) then were the first to apply these terms to overcomplete ICA, which was further studied and applied by Lee et al. Lee et al. (2000). De Lathauwer et al. Lathauwer et al. (1999) 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 Bofill (2000) 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).

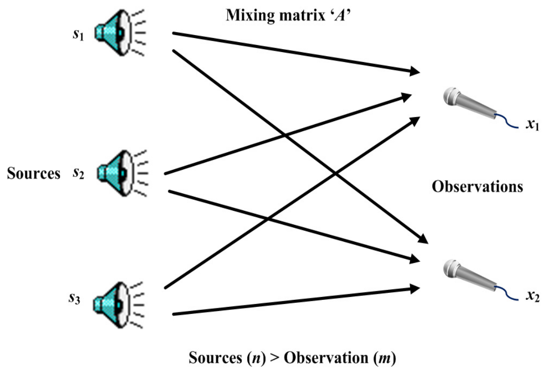

Fig. 9. Illustration of "overcomplete ICA"

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):

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

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.

2.2 Undercomplete ICA

The mixture of unknown sources is referred to as under-complete 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):

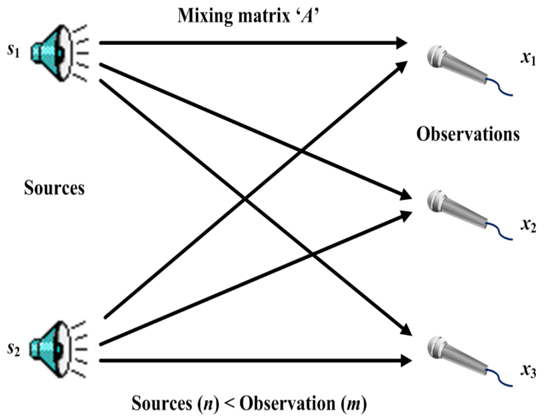

Fig. 10. Illustration of "undercomplete ICA"

$$\begin{aligned} x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) \\ x_2(t) &= a_{21}s_1(t) + a_{22}s_2(t) \\ x_3(t) &= a_{31}s_1(t) + a_{32}s_2(t) \end{aligned} \tag{24}$$

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.

3. Applications of ICA

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

- Machine fault detection Kano et al. (2003); Li et al. (2006); Ypma et al. (1999); Zhonghai et al. (2009)
- Seismic monitoring Acernese et al. (2004); de La et al. (2004)
- Reflection canceling Farid & Adelson (1999); Yamazaki et al. (2006)
- Finding hidden factors in financial data Cha & Chan (2000); Coli et al. (2005); Wu & Yu (2005)
- Text document analysis Bingham et al. (2002); Kolenda (2000); Pu & Yang (2006)
- Radio communications Cristescu et al. (2000); Huang & Mar (2004)
- Audio signal processing Cichocki & Amari (2002); Lee (1998)
- Image processing Cichocki & Amari (2002); Déniz et al. (2003); Fiori (2003); Karoui et al. (2009); Wang et al. (2008); Xiaochun & Jing (2004); Zhang et al. (2007)
- Data mining Lee et al. (2009)
- Time series forecasting Lu et al. (2009)
- Defect detection in patterned display surfaces Lu1 & Tsai (2008); Tsai et al. (2006)
- Bio medical signal processing Azzerboni et al. (2004); Castells et al. (2005); De Martino et al. (2007); Enderle et al. (2005); James & Hesse (2005); Kumagai & Utsugi (2004); Llinares & Igual (2009); Safavi et al. (2008); Zhu et al. (2006).

3.1 Audio and 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 Niedermeyer & Da Silva (1999); Rajapakse et al. (2002).
- Enhancement of low level Electrocardiogram components Niedermeyer & Da Silva (1999); Rajapakse et al. (2002)
- Separation of transplanted heart signals from residual original heart signals Wisbeck et al. (1998)
- Separation of low level myoelectric muscle activities to identify various gestures Calinon & Billard (2005); Kato et al. (2006); Naik et al. (2006; 2007)

One successful and promising application domain of blind signal processing includes those biomedical signals acquired using multi-electrode devices: Electrocardiography (ECG), Llinares & Igual (2009); Niedermeyer & Da Silva (1999); Oster et al. (2009); Phlypo et al. (2007); Rajapakse et al. (2002); Scherg & Von Cramon (1985); Wisbeck et al. (1998), Electroencephalography (EEG) Jervis et al. (2007); Niedermeyer & Da Silva (1999); Onton et al. (2006); Rajapakse et al. (2002); Vigário et al. (2000); Wisbeck et al. (1998), Magnetoencephalography (MEG) Hämäläinen et al. (1993); Mosher et al. (1992); Parra et al. (2004); Petersen et al. (2000); Tang & Pearlmutter (2003); Vigário et al. (2000).

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 Bell & Sejnowski (1997); Hyvarinen et al. (2001).

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) Hyvarinen et al. (2001); Lee (1998).

4. Conclusions

This chapter 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. The application domains of this novel technique are presented. The material covered in this chapter 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.

The other novel and recent advances of ICA, especially on Audio and Biosignal topics are covered in rest of the chapters in this book.

5. References

Acernese, F., Ciaramella, A., De Martino, S., Falanga, M., Godano, C. & Tagliaferri, R. (2004). Polarisation analysis of the independent components of low frequency events at stromboli volcano (eolian islands, italy), *Journal of Volcanology and Geothermal Research* 137(1-3): 153–168.

Attias, H. & Schreiner, C. E. (1998). Blind source separation and deconvolution: the dynamic component analysis algorithm, *Neural Comput.* 10(6): 1373–1424.

Azzerboni, B., Carpentieri, M., La Foresta, F. & Morabito, F. C. (2004). Neural-ica and wavelet transform for artifacts removal in surface emg, *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, Vol. 4, pp. 3223–3228 vol.4.

Bell, A. J. & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution., *Neural Comput* 7(6): 1129–1159.

Bell, A. J. & Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters., *Vision Res* 37(23): 3327–3338.

Bingham, E., Kuusisto, J. & Lagus, K. (2002). Ica and som in text document analysis, *SIGIR '02: Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM, pp. 361–362.

Bofill (2000). Blind separation of more sources than mixtures using sparsity of their short-time fourier transform, pp. 87–92.

Calinon, S. & Billard, A. (2005). Recognition and reproduction of gestures using a probabilistic framework combining pca, ica and hmm, *ICML '05: Proceedings of the 22nd international conference on Machine learning*, ACM, pp. 105–112.

Castells, F., Igual, J., Millet, J. & Rieta, J. J. (2005). Atrial activity extraction from atrial fibrillation episodes based on maximum likelihood source separation, *Signal Process.* 85(3): 523–535.

Cha, S.-M. & Chan, L.-W. (2000). Applying independent component analysis to factor model in finance, *IDEAL '00: Proceedings of the Second International Conference on Intelligent Data Engineering and Automated Learning, Data Mining, Financial Engineering, and Intelligent Agents*, Springer-Verlag, pp. 538–544.

Cichocki, A. & Amari, S.-I. (2002). *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, John Wiley & Sons, Inc.

Coli, M., Di Nisio, R. & Ippoliti, L. (2005). Exploratory analysis of financial time series using independent component analysis, *Information Technology Interfaces, 2005. 27th International Conference on*, pp. 169–174.

Cristescu, R., Ristaniemi, T., Joutsensalo, J. & Karhunen, J. (2000). Cdma delay estimation using fast ica algorithm, Vol. 2, pp. 1117–1120 vol.2.

de La, Puntonet, C. G., Górriz, J. M. & Lloret, I. (2004). An application of ica to identify vibratory low-level signals generated by termites, pp. 1126–1133.

De Martino, F., Gentile, F., Esposito, F., Balsi, M., Di Salle, F., Goebel, R. & Formisano, E. (2007). Classification of fmri independent components using ic-fingerprints and support vector machine classifiers, *NeuroImage* 34: 177–194.

Déniz, O., Castrillón, M. & Hernández, M. (2003). Face recognition using independent component analysis and support vector machines, *Pattern Recogn. Lett.* 24(13): 2153–2157.

Enderle, J., Blanchard, S. M. & Bronzino, J. (eds) (2005). *Introduction to Biomedical Engineering, Second Edition*, Academic Press.

Farid, H. & Adelson, E. H. (1999). Separating reflections and lighting using independent components analysis, *cvpr* 01.

Fiori, S. (2003). Overview of independent component analysis technique with an application to synthetic aperture radar (sar) imagery processing, *Neural Netw.* 16(3-4): 453–467.

Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J. & Lounasmaa, O. V. (1993). Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain, *Reviews of Modern Physics* 65(2): 413+.

Hansen (2000). *Blind separation of noicy image mixtures.*, Springer-Verlag, pp. 159–179.

Harpur, G. F. & Prager, R. W. (1996). Development of low entropy coding in a recurrent network., *Network (Bristol, England)* 7(2): 277–284.

Herault, J. & Jutten, C. (1987). Herault, J. and Jutten, C. (1987), Space or time adaptive signal processing by neural network models, in 'AIP Conference Proceedings 151 on Neural Networks for Computing', American Institute of Physics Inc., pp. 206-211.

Huang, J. P. & Mar, J. (2004). Combined ica and fca schemes for a hierarchical network, *Wirel. Pers. Commun.* 28(1): 35–58.

Hyvarinen, A., Cristescu, R. & Oja, E. (1999). A fast algorithm for estimating overcomplete ica bases for image windows, *Neural Networks, 1999. IJCNN '99. International Joint Conference on*, Vol. 2, pp. 894–899 vol.2.

Hyvarinen, A., Karhunen, J. & Oja, E. (2001). *Independent Component Analysis*, Wiley-Interscience.

James, C. J. & Hesse, C. W. (2005). Independent component analysis for biomedical signals, *Physiological Measurement* 26(1): R15+.

Jervis, B., Belal, S., Camilleri, K., Cassar, T., Bigan, C., Linden, D. E. J., Michalopoulos, K., Zervakis, M., Besleaga, M., Fabri, S. & Muscat, J. (2007). The independent components of auditory p300 and cnv evoked potentials derived from single-trial recordings, *Physiological Measurement* 28(8): 745–771.

Kab'an (2000). Clustering of text documents by skewness maximization, pp. 435–440.

Kano, M., Tanaka, S., Hasebe, S., Hashimoto, I. & Ohno, H. (2003). Monitoring independent components for fault detection, *AIChE Journal* 49(4): 969–976.

Karoui, M. S., Deville, Y., Hosseini, S., Ouamri, A. & Ducrot, D. (2009). Improvement of remote sensing multispectral image classification by using independent component analysis, *2009 First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, IEEE, pp. 1–4.

Kato, M., Chen, Y.-W. & Xu, G. (2006). Articulated hand tracking by pca-ica approach, *FGR '06: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, IEEE Computer Society, pp. 329–334.

Kolenda (2000). *Independent components in text*, Advances in Independent Component Analysis, Springer-Verlag, pp. 229–250.

Kumagai, T. & Utsugi, A. (2004). Removal of artifacts and fluctuations from meg data by clustering methods, *Neurocomputing* 62: 153–160.

Lathauwer, D., L. Comon, P., De Moor, B. & Vandewalle, J. (1999). Ica algorithms for 3 sources and 2 sensors, *Higher-Order Statistics, 1999. Proceedings of the IEEE Signal Processing Workshop on*, pp. 116–120.

Lee, J.-H. H., Oh, S., Jolesz, F. A., Park, H. & Yoo, S.-S. S. (2009). Application of independent component analysis for the data mining of simultaneous eeg-fmri: preliminary experience on sleep onset., *The International journal of neuroscience* 119(8): 1118–1136. URL: *http://view.ncbi.nlm.nih.gov/pubmed/19922343*

Lee, T. W. (1997). Blind separation of delayed and convolved sources, pp. 758–764.

Lee, T. W. (1998). *Independent component analysis: theory and applications*, Kluwer Academic Publishers.

Lee, T. W., Girolami, M., Lewicki, M. S. & Sejnowski, T. J. (2000). Blind source separation of more sources than mixtures using overcomplete representations, *Signal Processing Letters, IEEE* 6(4): 87–90.

Lee, T. W., Lewicki, M. S. & Sejnowski, T. J. (1999). Unsupervised classification with non-gaussian mixture models using ica, *Proceedings of the 1998 conference on Advances in neural information processing systems*, MIT Press, Cambridge, MA, USA, pp. 508–514.

Lewicki, M. S. & Sejnowski, T. J. (2000). Learning overcomplete representations., *Neural Comput* 12(2): 337–365.

Li, Z., He, Y., Chu, F., Han, J. & Hao, W. (2006). 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* 291(1-2): 60–71.

Llinares, R. & Igual, J. (2009). Application of constrained independent component analysis algorithms in electrocardiogram arrhythmias, *Artif. Intell. Med.* 47(2): 121–133.

Lu, C.-J., Lee, T.-S. & Chiu, C.-C. (2009). Financial time series forecasting using independent component analysis and support vector regression, *Decis. Support Syst.* 47(2): 115–125.

Lu1, C.-J. & Tsai, D.-M. (2008). Independent component analysis-based defect detection in patterned liquid crystal display surfaces, *Image Vision Comput.* 26(7): 955–970.

Mackay, D. J. C. (1996). Maximum likelihood and covariant algorithms for independent component analysis, *Technical report*, University of Cambridge, London.

Mosher, J. C., Lewis, P. S. & Leahy, R. M. (1992). Multiple dipole modeling and localization from spatio-temporal meg data, *Biomedical Engineering, IEEE Transactions on* 39(6): 541–557.

Naik, G. R., Kumar, D. K., Singh, V. P. & Palaniswami, M. (2006). Hand gestures for hci using ica of emg, *VisHCI '06: Proceedings of the HCSNet workshop on Use of vision in human-computer interaction*, Australian Computer Society, Inc., pp. 67–72.

Naik, G. R., Kumar, D. K., Weghorn, H. & Palaniswami, M. (2007). Subtle hand gesture identification for hci using temporal decorrelation source separation bss of surface emg, *Digital Image Computing Techniques and Applications, 9th Biennial Conference of the Australian Pattern Recognition Society on*, pp. 30–37.

Niedermeyer, E. & Da Silva, F. L. (1999). *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, Lippincott Williams and Wilkins; 4th edition .

Olshausen (1995). Sparse coding of natural images produces localized, oriented, bandpass receptive fields, *Technical report*, Department of Psychology, Cornell University.

Olshausen, B. A. (1996). Learning linear, sparse, factorial codes, *Technical report*.

Onton, J., Westerfield, M., Townsend, J. & Makeig, S. (2006). Imaging human eeg dynamics using independent component analysis, *Neuroscience & Biobehavioral Reviews* 30(6): 808–822.

Oster, J., Pietquin, O., Abächerli, R., Kraemer, M. & Felblinger, J. (2009). Independent component analysis-based artefact reduction: application to the electrocardiogram for improved magnetic resonance imaging triggering, *Physiological Measurement* 30(12): 1381–1397.
URL: *http://dx.doi.org/10.1088/0967-3334/30/12/007*

Parra, J., Kalitzin, S. N. & Lopes (2004). Magnetoencephalography: an investigational tool or a routine clinical technique?, *Epilepsy & Behavior* 5(3): 277–285.

Petersen, K., Hansen, L. K., Kolenda, T. & Rostrup, E. (2000). On the independent components of functional neuroimages, *Third International Conference on Independent Component Analysis and Blind Source Separation*, pp. 615–620.

Phlypo, R., Zarzoso, V., Comon, P., D'Asseler, Y. & Lemahieu, I. (2007). Extraction of atrial activity from the ecg by spectrally constrained ica based on kurtosis sign, *ICA'07: Proceedings of the 7th international conference on Independent component analysis and signal separation*, Springer-Verlag, Berlin, Heidelberg, pp. 641–648.

Pu, Q. & Yang, G.-W. (2006). Short-text classification based on ica and lsa, *Advances in Neural Networks - ISNN 2006* pp. 265–270.

Rajapakse, J. C., Cichocki, A. & Sanchez (2002). Independent component analysis and beyond in brain imaging: Eeg, meg, fmri, and pet, *Neural Information Processing, 2002. ICONIP '02. Proceedings of the 9th International Conference on*, Vol. 1, pp. 404–412 vol.1.

Safavi, H., Correa, N., Xiong, W., Roy, A., Adali, T., Korostyshevskiy, V. R., Whisnant, C. C. & Seillier-Moiseiwitsch, F. (2008). Independent component analysis of 2-d electrophoresis gels, *ELECTROPHORESIS* 29(19): 4017–4026.

Scherg, M. & Von Cramon, D. (1985). Two bilateral sources of the late aep as identified by a spatio-temporal dipole model., *Electroencephalogr Clin Neurophysiol* 62(1): 32–44.

Sorenson (2002). Mean field approaches to independent component analysis, *Neural Computation* 14: 889–918.

Stone, J. V. (2004). *Independent Component Analysis : A Tutorial Introduction (Bradford Books)*, The MIT Press.

Tang, A. C. & Pearlmutter, B. A. (2003). Independent components of magnetoencephalography: localization, pp. 129–162.

Tsai, D.-M., Lin, P.-C. & Lu, C.-J. (2006). An independent component analysis-based filter design for defect detection in low-contrast surface images, *Pattern Recogn.* 39(9): 1679–1694.

Vigário, R., Särelä, J., Jousmäki, V., Hämäläinen, M. & Oja, E. (2000). Independent component approach to the analysis of eeg and meg recordings., *IEEE transactions on bio-medical engineering* 47(5): 589–593.

Wang, H., Pi, Y., Liu, G. & Chen, H. (2008). Applications of ica for the enhancement and classification of polarimetric sar images, *Int. J. Remote Sens.* 29(6): 1649–1663.

Wisbeck, J., Barros, A. & Ojeda, R. (1998). Application of ica in the separation of breathing artifacts in ecg signals.

Wu, E. H. & Yu, P. L. (2005). Independent component analysis for clustering multivariate time series data, pp. 474–482.

Xiaochun, L. & Jing, C. (2004). An algorithm of image fusion based on ica and change detection, *Proceedings 7th International Conference on Signal Processing, 2004. Proceedings. ICSP '04. 2004.*, IEEE, pp. 1096–1098.

Yamazaki, M., Chen, Y.-W. & Xu, G. (2006). Separating reflections from images using kernel independent component analysis, *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, Vol. 3, pp. 194–197.

Ypma, A., Tax, D. M. J. & Duin, R. P. W. (1999). Robust machine fault detection with independent component analysis and support vector data description, *Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop*, pp. 67–76.

Zhang, Q., Sun, J., Liu, J. & Sun, X. (2007). A novel ica-based image/video processing method, pp. 836–842.

Zhonghai, L., Yan, Z., Liying, J. & Xiaoguang, Q. (2009). Application of independent component analysis to the aero-engine fault diagnosis, *2009 Chinese Control and Decision Conference*, IEEE, pp. 5330–5333.

Zhu, Y., Chen, T. L., Zhang, W., Jung, T.-P., Duann, J.-R., Makeig, S. & Cheng, C.-K. (2006). Noninvasive study of the human heart using independent component analysis, *BIBE '06: Proceedings of the Sixth IEEE Symposium on BionInformatics and BioEngineering*, IEEE Computer Society, pp. 340–347.